

GenCore version 5.1.6  
Copyright (c) 1993 - 2004 Compugen Ltd.

OM protein - protein search, using SW model

Run on: February 5, 2004, 13:18:44 ; Search time 46 Seconds  
(without alignments)  
750.534 Million cell updates/sec

Title: US-09-972-758a-2

Sequence: 1 MAEPFLSEYQHQPQTSNCTG.....LTENELHROQERAPLSKFD 359

Scoring table: BLOSUM62  
Gapop 10.0 , Gapext 0.5

Searched: 283308 seqs, 9616682 residues

Total number of hits satisfying chosen parameters: 283308

Minimum DB seq length: 0  
Maximum DB seq length: 200000000

Post-processing: Minimum Match 0%  
Maximum Match 100%

Listing first 45 summaries

Database :  
1: p1r1:\*  
2: p1r2:\*  
3: p1r3:\*  
4: p1r4:\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

## SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	148.5	7.8	992	2 T46337	hypothetical prote
2	139	7.3	755	2 S32103	filensin - bovine
3	134	7.0	568	1 I61106	involucrin - rat
4	133	7.0	462	1 A60746	chromogranin A pre
5	133	7.0	538	1 FOMVIM	gag polyprotein
6	133	7.0	1493	2 A44224	DNA repair helicase
7	132.5	6.9	530	2 TC7168	lens epithelium-de
8	131	6.9	681	2 ES9436	KIMA1314 protein
9	131	6.9	1095	2 T00329	hypothetical prote
10	131	6.9	1898	1 A45973	trichoglycin - hum
11	131	6.9	3488	2 T34418	hypothetical prote
12	129	6.8	1054	2 T30177	cytoskeleton assem
13	128	6.7	1733	1 B45344	probable nuclear a
14	127.5	6.7	852	2 T06310	hypothetical prote
15	127.5	6.7	918	2 A88188	plectin - rat
16	127.5	6.7	4687	1 A39638	protein kinase (EC
17	126.5	6.6	777	2 F54024	trichoglycin - rab
18	126.5	6.6	1407	1 S28589	protein kinase (EC
19	125.5	6.6	779	2 ES4024	trichoglycin - she
20	125.5	6.6	1549	1 A40681	trichoglycin - she
21	125.5	6.6	1909	2 A45592	liver stage antige
22	125	6.5	383	2 C96581	hypothetical prote
23	125	6.5	913	2 TS2485	neurofilament prot
24	124.5	6.5	474	2 T16441	hypothetical prote
25	123.5	6.5	677	1 S09078	chromogranin B pre
26	123.5	6.5	693	2 JC7925	nucleolin - common
27	123.5	6.5	1870	2 S37671	MHC class III hist
28	123.5	6.5	1872	2 S36152	MHC class III hist
29	123	6.4	723	2 A48217	single-strand DNA/

30	123	6.4	1829	2 T41751	1-afadin - rat
31	122	6.4	705	2 S32644	nucleolin - Africa
32	122	6.4	723	2 S33688	hypothetical prote
33	121.5	6.4	1974	2 T30010	hypothetical prote
34	121	6.3	538	1 FOMVIM	gag polyprotein -
35	120.5	6.3	849	1 S64732	scaffold attachmen
36	120	6.3	536	1 FOMVIM	gag polyprotein -
37	120	6.3	699	2 I38073	nucleolar phosphop
38	120	6.3	1325	2 S16129	dynein-associated
39	120	6.3	1663	2 T42092	s-afadin - rat
40	119.5	6.3	451	2 G70241	hypothetical prote
41	119.5	6.3	651	2 F86563	hypothetical prote
42	119.5	6.3	651	2 A72060	hypothetical prote
43	119.5	6.3	1233	2 T14157	serine/threonine p
44	119.5	6.3	1560	2 T30282	calcium-binding pr
45	119	6.2	449	1 A41520	chromogranin A pre

## ALIGNMENTS

RESULT 1		T46337		hypothetical protein DKFZp43402413.1 - human (fragment)	
C/Species: Homo sapiens (man)		C/Date: 04-Feb-2000		#sequence_revision 04-Feb-2000 #text_change 04-Feb-2000	
Accession: T46337		Submitted to the Protein Sequence Database, January 2000			
A/Reference number: Z23037		A/Status: preliminary			
A/Accession: T46337		A/Molecule type: mRNA			
A/Residues: 1992 <AAA>		A/Cross-References: EMBL:AL137265			
A/Experimental source: adult testis; clone DKFZp43402413		C/Genetics:			
A/Note: DKFZp43402413.1					
Query Match		7.8%; Score 148.5; DB 2; Length 992;			
Best Local Similarity 21.1%; Pred. No. 0.19;					
Matches 94; Conservative 66; Mismatches 149; Indels 137; Gaps 17;					
QY	25	OEEINERPPGAEERVP-----EEDSRWQSRAP-----POLGGRGPGGE-----	64		
DB	230	QDELOSKQKGLERLSPPLPHERRAQSPRLATMEERPQ-----GPGQPEWKAEEL	284		
QY	65	-----GSLESQPPPIQTA-CPESSCLREGEKQNGDSSAGDPPPAE---VEPTP-	113		
DB	285	GEDSAASLSLQSLQREQAPEPPACCEKQEQSHQAEILGPGQEADEBEKVAVSPTP	344		
QY	114	-----EAEILAPQPCDSASKLGAPAGGEEWQQQQLGKK-----	152		
DB	345	VSEVSTPEVPAPPEQLSBA--LKAMEAVAVQLTEDQHLLESQEKQQLREKLCOE	403		
QY	153	-----HRRRPSKKRHKPYKLTWEKKKFEDEKQSRASIRAEWPAKQVPAYNT	205		
DB	404	EEETLRHQKEQSLSLERLQKALEEBAKREESQRLSLRNQVOSTQA-----	457		
QY	206	TOFLMDHOOEEDLTGGLYSKRAAKSD-----DTSDDPFMEGGEEDGS-----	252		
DB	458	-----DEQIRAEQASLQKLRBELSQKARASLEQKNRMQLKEIRASKESEQ	511		
QY	253	-----DMGDDGEFLQRPDSFYERY-----HTE---SLQMSK	284		
DB	512	AALNAKAKALQQLREQLGERKEAVATLEKHSABLERLCSLEAKREVVSSLSKQIQ	571		
QY	285	OELIKETLEKCLSMEDENNRRLRESKRLGGDDRVREL-----ELTDRU	332		
DB	572	EAQKEEAQQLQKLGVE---HVVQKSYHVAGYHEHLSLREKQAEVGEHERLDM	628		
QY	333	RAENTQLLT-----ENELHROQERAPL	354		



R,hutton, J.C.; Nielsen, E.; Kasten, W.  
 FEBS Lett. 236, 269-274, 1988  
 A>Title: The molecular cloning of the chromogranin A-like precursor of beta-granin and B  
 A/Reference number: S02543; MUID:88312980; PMID:3044825  
 A/Accession: S02543  
 A>Status: nucleic acid sequence not shown; not compared with conceptual translation  
 A/Molecule type: mRNA  
 A/Residues: 19-107, 'QQQQ', 108-462 <HDT>  
 A/Note: part of this sequence, including the amino end of beta-granin, was determined by  
 C/Superfamily: chromogranin A  
 C/Keywords: glycoprotein; phosphoprotein  
 F/1-18/Domain: signal sequence #status predicted <SIG>  
 F/19-462/Product: chromogranin A #status predicted <WAT>  
 F/92-107/Region: glutamine-rich  
 F/417-419/Region: cell attachment (R-G-D) motif  
 F/35-56/Disulfide bonds: #status predicted  
 F/68-147,216,349,358,408,409,413/Binding site: phosphate (Ser) (covalent) #status pr  
 F/189/Binding site: carbohydrate (Asn) (covalent) #status predicted

Query Match 7.0%; Score 133; DB 1; Length 462;  
 Best Local Similarity 24.2%; Pred. No. 0.7;  
 Matches 87; Conservative 54; Mismatches 159; Indels 60; Gaps 17;

```

QY 3 EPLLEYQHPQPSNCTGAAGVDELNPPRPPGAEERVPEDSRWQSRAPFQGLGRPP 62
DB 64 ERYLSILRHQ-----NLKEIQDLALQGAKEAQQQQQQQQQQ-QQQQOHSSFE 112
QY 63 GEGS--LESGPPPLQ-----TQACPESSCLREGEGQNGDSSAGG--DPE-PPAEVEPT 113
DB 113 DELSEVFENQSPAKHGDAASAPSKDTVEREDSDKQGDAPETGTEGPPAPFPBPQ 172
QY 114 EAELL--AQCCHDSSEASKLAPAGGEWQGOQROLGKKHRRPS--KKKHMKPYRK 169
DB 173 ESSMNGNSQSGPBDTANNTPSPISLPSQEHGIPQTGSE--RGPSAQQAARAKQEEK 229
QY 170 LTWEKKKFKDEKQSLRASRIAPENFAKGPVAPYNT-----TQFLMDHDQEBPLKTG 223
DB 230 EEEEEEKEEEEEEKEEVAIRERAKGPEVPTAASSSHFYGYKKIQKDDQSGESQAVNG 289
QY 224 LYSKRAAKSDDTSDDFMEGEGEDGSGDM-----GGDGSFLLORDSETYERY 274
DB 290 --KTGSAVSPBEGKGLHSSQEEBEGEAMGPPGGLPPGGQGS--LERKQEEEE-- 343
QY 275 HTESLQMSKQELIKELYLEKCLSRMEDENNRRLRESKRLGGDDARVLELELDRLRA 344
DB 344 -----EEERLSREM--EDKMSKMDQAKELTAE--KLEGEEDDPRSMKLSF--RARA 391

```

RESULT 5  
 FOMVIM  
 gag polypeptide - Moloney murine leukemia virus  
 N/containe: core protein p15; core shell protein p30; inner coat protein p12; nucleoprot  
 C/Species: Moloney murine leukemia virus  
 A/Note: host Mus sp. (mouse)  
 C/Date: 01-Sep-1981 #sequence\_revision 27-Nov-1985 #text\_change 24-Jul-1997  
 C/Accession: A03930  
 R/Shinnick, T.M.; Lerner, R.A.; Sutcliffe, J.G.  
 Nature 293, 543-548, 1981  
 A>Title: Nucleotide sequence of Moloney murine leukaemia virus.  
 A/Reference number: A93265; MUID:82035843; PMID:6169994  
 A/Accession: A03930  
 A/Molecule type: genomic RNA  
 A/Residues: 1-538 <SHI>  
 A/Experimental source: clone pMLV-1  
 C/Comment: This protein is synthesized as a gag-pol polypeptide.  
 C/Genetics:  
 A/Genes: gag  
 C/Superfamily: mammalian retrovirus gag polypeptide 1  
 C/Keywords: core protein, inner coat protein, nucleoprotein, polypeptide  
 F/2-133/Product: core protein p15 #status predicted <CIS>  
 F/132-215/Product: inner coat protein p12 #status predicted <C12>  
 F/216-478/Product: core shell protein p30 #status predicted <C30>  
 F/479-534/Product: nucleoprotein p10 #status predicted <C10>

Query Match 7.0%; Score 133; DB 1; Length 538;  
 Best Local Similarity 20.9%; Pred. No. 0.82;  
 Matches 89; Conservative 40; Mismatches 142; Indels 154; Gaps 19;

```

QY 30 PERPPA-----EERVVEDSRWQSRAPFQGLGRPPG---GEES-----LESGPP 74
DB 107 PPLPAPSLPLEPPRSTPRSSLYPALP--PSLGAKPKQVLSDSGLIDLTEDPPY 165
QY 75 QTOACPESSCLREGEGQNGDSSAGDPPPAEVEPTPEALL-----AQCCHDSASKL 130
DB 166 RDRPPPSD--RDG-----NGERTPAGAP-----DPSMASRLRRRRPPVADSTTSSQA 214
QY 131 GAPAGGEWQGOQROLGKKHRRPSKKKHMKPYRKLTWEKKKFKDEKQSLASRI 190
DB 215 PFLRAGN-----GQLQWPFSSSDLYNWK-----NNNSFSEDPKLTALIE 257
QY 191 AEMFAKGPVAPYNTTQFLMDHDQEBPLKTG-----LYSKRAAKSD-- 235
DB 258 SVLI-----THQPTWDDCOQLGLTLTGEEKQVLEARKAVRDDGPRPTQLPNE 307
QY 236 -----TSDDFMEGEE-----DGSDMGDGSSE 261
DB 308 VDAAPFLERPPDDYTTQAGRHLVHYRQLLAGLQVAGSPNTLAKVKITQGPNSPBA 367
QY 262 FLORDSETYERY-----HTESLQMSKQELIKE 290
DB 368 FLER--LKEAVRRRTVPDPEDPQETVNSMFTWQSAFDIGRLGREDLKNNTLGDVLR-- 425
QY 291 YLELEKCLSRMEDENNRRLRESKRLGGDDARV--RELELDRLRAENLQTLTENELRHQ 349
DB 426 --EAEKIFPKRETPEER-----EERIRRETEKEEBRRRTDEQEKERDRRRR 472
QY 350 ERAPL 354
DB 473 EWSKL 477

```

RESULT 6  
 A44224  
 DNA repair helicase ERCC6 - human  
 C/Species: Homo sapiens (man)  
 C/Date: 10-Jun-1993 #sequence\_revision 18-Nov-1994 #text\_change 21-Jul-2000  
 C/Accession: A44224  
 R/Troelstra, C.; van Gool, A.; de Wit, J.; Vermeulen, W.; Bootsma, D.; Hoeijmakers, J.H.  
 Cell 71, 939-953, 1992  
 A>Title: ERCC6, a member of a subfamily of putative helicases, is involved in Cockayne's  
 A/Reference number: A44224; MUID:93092214; PMID:1339317  
 A/Accession: A44224  
 A/Status: preliminary; not compared with conceptual translation  
 A/Molecule type: nucleic acid  
 A/Residues: 1-1493 <TRO>  
 A/Cross-references: GB:L04791; NID:g182180; PIDN:AA52397.1; PID:g182181  
 A/Note: sequence extracted from NCBI backbone (NCBIR:120238)  
 C/Keywords: DNA binding

Query Match 7.0%; Score 133; DB 2; Length 1493;  
 Best Local Similarity 23.7%; Pred. No. 2.3; Indels 104; Gaps 17;  
 Matches 82; Conservative 42; Mismatches 118; Indels 104; Gaps 17;

```

QY 35 GAERVP-----EEDSRWQSRAPFQGLGRPGEGSLGSPPLQTOACPESSCLREG 88
DB 201 GAEVKIELDHASLEEDA-----EPGP--SSLSGMLMPVQETAME--LIRTG 243
QY 89 EKGQNGDSSAGDPPPAEVEPTP--EA-----ELLAQCCHDSSEASK 129
DB 244 QM-----TPFGTQIFQKQEKPRKIMLEASGFERYTLDOAKLSFERKKQCGCNKRAAK 297
QY 130 LGAPA-----AGGEWQGOQROLGKKHRRPSKKKHMKPYRKLTWEKKKFKDEKQSL 184
DB 298 AAPVPTPPAPVONKPNKARLASKKER-----LKKHITKLQKRALQFQGVLPKAR 352
QY 185 RA-----SRIAEWFAKGPVAPYNTTQFLMDHDQEBPLD-----KTG 223

```

Db 353 RPMSDMPREAEDEGESE-SEVPTHEEEEDDEVEGAALSGDGTDELKPLPKG 411  
 QY 224 LYSRAAAKSDTDDFMEEGEE-DGSGDGMGDSSEFLQDPFSEYRYHTESLQNM 282  
 Db 412 KRQKVPVQV---IDDDFFSSGEAEAAVSGEGGGRVGVRRDDGDDYKQRLRW 468  
 QY 283 SKQELIKYLELEKLSRMEDENNRLLESKRLLGGDDARVLELE 328  
 Db 469 NK-----LRLQDKERLKLFE-----DDSESDAEFD 494

## RESULT 7

UC7168  
 lens epithelium-derived growth factor - human  
 N:Alternate names: transcriptional co-activator p75 protein  
 C:Species: Homo sapiens (man)  
 C>Date: 04-Mar-2000 #sequence\_revision 04-Mar-2000 #text\_change 11-May-2000  
 C/Accession: J07168  
 R/Singh, D.P.; Ohguro, N.; Kikuchi, T.; Sueno, T.; Reddy, V.N.; Yuge, K.; Chylack Jr., I.  
 Biochem. Biophys. Res. Commun. 267, 373-381, 2000  
 A/Title: Lens epithelium-derived growth factor: Effects on growth and survival of lens  
 A/Reference number: J07168; MUID:20090645; PMID:10623627  
 A/Accession: J07168  
 A/Molecule type: mRNA  
 A/Residues: 1-530 <SIN>  
 A/Cross-references: GB:AF063020  
 A/Experimental source: lens epithelial cells  
 C/Comment: This protein, a regulatory factor, belongs to a family of growth factor, which  
 on of this factor stimulates cell attachment, shortens the initial lag in growth, and  
 C/Genetics:  
 A:Gene: ledgf  
 C/Keywords: eye lens; glycoprotein; growth factor; growth regulation

Query Match 6.9%; Score 132.5; DB 2; Length 530;  
 Best Local Similarity 21.2%; Pred. No. 0.87;  
 Matches 77; Conservative 56; Mismatches 115; Indels 115; Gaps 17;  
 QY 7 SEYQHPOTSNTGAAVAOELNPERPPGAERVPEDSRQSAFPQ-----L 55  
 Db 93 SSQAATKGSNMSADVBEK-----ETSVKEDTHEKASNEVDTKAVDITTK 143  
 QY 56 GGRPG--PEGEGSLESQPPPLQTOACPESSCLREGEKGNGDSSAGDPPEAEV--E 110  
 Db 144 AARGRRKRAKQVETEBAGVTTATASVULKVSPKGR-----PAATEVKIRK 192  
 QY 111 PTBEALLQPCDSEASKUGAPAGGEEMGQQQLGKKRRRRPKKKRMKPYKL 170  
 Db 193 PRGRPKVXQPC-PSSESDI-----TEEDSKSKKKQKEKQPKK----- 229  
 QY 171 TWEEKKKFDEKQSLRASIRAEWFAKQPVAPYNTTOFLMDHDOEEDLKTG--LYSK 227  
 Db 230 ---QPKDEEGQK-----EDPRKEPDKKEGKEVESEK 260  
 QY 228 RA-AAKSDTSDDDFMEEGEEEDGSDMGDSSEFLQDPFSEYRYHTESLQ 280  
 Db 261 RKULAKTVGTVSTDSSEEGDQGEKKRKG-----RNFOAHRNRRLKQGEKEAD 313  
 QY 281 NMSKQELIKYLELEKLSRMEDENNR--LLESKRLLGGDDARVLELEL-DRLRAEN 336  
 Db 314 RKRKQ---EOMTEQ---ONKDEGKPEVKVKEKRETSMDSRIRIHAENKSLKIN 367  
 QY 337 LQL 339  
 Db 368 LDV 370

## RESULT 8

ES9436  
 KIAA1314 protein (imported) - human  
 C/Species: Homo sapiens (man)  
 C>Date: 03-Jun-2002 #sequence\_revision 03-Jun-2002 #text\_change 23-Sep-2002  
 C/Accession: E59436; F59436

R/Nagase, T.; Kikuno, R.; Ishikawa, K.I.; Hirosewa, M.; Ohara, O.  
 DNA Res. 7, 65-73, 2000  
 A/Title: Prediction of the coding sequences of unidentified human genes. XVI. The complet  
 A/Reference number: E59436  
 A/Accession: E59436  
 A/Status: preliminary  
 A/Molecule type: mRNA  
 A/Residues: 1-681 <NAG>  
 A/Cross-references: GB:BA92552; PID:g7243009; PIDN:BA92552.1  
 R/Ohara, O.; Nagase, T.; Kikuno, R.  
 submitted to GenBank, January 2000  
 A/Reference number: F59436  
 A/Accession: F59436  
 A/Status: preliminary  
 A/Molecule type: mRNA  
 A/Residues: 1-681 <NAG>  
 A/Cross-references: GB:BA92552; PID:g7243009; PIDN:BA92552.1

Query Match 6.9%; Score 131; DB 2; Length 681;  
 Best Local Similarity 26.4%; Pred. No. 1.4;  
 Matches 84; Conservative 44; Mismatches 110; Indels 80; Gaps 18;

QY 66 SLESQPPPIQTOACPESSC-----LREGEKGNGDSSAGG--DFPPAEVEPPEAE 116  
 Db 14 SNESLHPAPFSRNSASVDSASMEDPWRKESIK--DSMGQGEPPPAEVTVPDGE 70  
 QY 117 LLAQPCDSEASKUGAPAGGEEMGQQQLGKKRRRRPKKKRMKPYKL 176  
 Db 71 LEAEMQDVGLSTL---ISGDEEDG--KALLSTLRTQAAVQK---YHTYQTMK 121  
 QY 177 KQDEKQSLASRIRAEWFAKQ---PVAPYNTTOFLMDHDOEEDL-----KTGLYSK 227  
 Db 122 K--DKQSIRDV---DIFGVSESPPRDTCGNHTNOL--DGKREBELPRVKTSGSMPD 173  
 QY 228 RAASDSDTSDDDFMEEGEEEDGSDMG-----GNGS-----BF 262  
 Db 174 DASLNTSLSDASQDEGSFAVPRSDVAILETIPVLPVHNSGSPPGQPVONALSDDF 233  
 QY 263 LQD-----FSEYRYHTESLQ--NMSKQELIKYLELEKLSRMEDENNRLRL- 310  
 Db 234 LKKNIPRAEELSFEVSVSEWTEALKKKSIIKE---DYVLTFRNVQKTRFGLT 269  
 QY 311 ESKRLGGDD-ARVLELEL 327  
 Db 290 EAGDLSAEDMKIRHLSL 307

## RESULT 9

T00329  
 hypothetical protein KIA0553 - human (fragment)  
 C/Species: Homo sapiens (man)  
 C>Date: 01-Feb-1999 #sequence\_revision 01-Feb-1999 #text\_change 21-Jul-2000  
 C/Accession: T00329  
 R/Nagase, T.; Ishikawa, K.; Miyajima, N.; Tanaka, A.; Kotani, H.; Nomura, N.; Ohara, O.  
 DNA Res. 5, 31-39, 1998  
 A/Title: Prediction of the coding sequences of unidentified human genes. IX. The complete  
 A/Reference number: Z14086; MUID:98290545; PMID:9628581  
 A/Accession: T00329  
 A/Status: preliminary; translated from GB/EMBL/DBJ  
 A/Molecule type: mRNA  
 A/Residues: 1-1095 <NAG>  
 A/Cross-references: EMBL:AB011125; NID:g3043629; PIDN:BA925479.1; PID:g3043630  
 A/Experimental source: brain  
 C/Genetics:  
 A/Note: KIA0553

Query Match 6.9%; Score 131; DB 2; Length 1095;  
 Best Local Similarity 19.7%; Pred. No. 2.2;  
 Matches 99; Conservative 41; Mismatches 137; Indels 226; Gaps 21;

QY 13 POTSNTGAAVAOELNPERPPGAERVPE--EDSRMGSRAFPQLGRRPGEG----- 65  
 Db 33 PRKSCIRAAASQ-----GAEKTVSEVSEQPKETSMTEP---SEPGSRAEAKKALG 80



QY 66 -----SLESQPPPL-OTQACPESSCLRE-----GEKGQNGDSSAGDPPP----- 105  
DB 81 GVDSDSLSHSQKVSSETQMC-ESNSKETSLATPAGKESQEGPKAPGTPFPFVLSSKDS 139  
QY 106 -----PAEV----- 109  
DB 140 TALQWSESLILITKAPPSISYSCNPLYPFKLSRNKDAFTKTEKPDIGSSSKDHLQGL 199  
QY 110 -----EPTPEALLAQPCHESEASKLGAAPAG-----E 138  
DB 200 DGEFHKSGVEGGEKTVRSRSGGMDAPAGSAGSLANKOEPGSHSGETEDTGRSLPSK 259  
QY 139 EEWGQOQOLQKKKRRRPSKKRRHWKPYKLTWEKK--KFEKQSLRASIR----- 190  
DB 260 ERSGKSHRKKKKKKKKKKKKH-----KADTEKSSVAESGEKSKKKRKKRKNKS 314  
QY 191 -----AEMFAKQOPVAPYNTTQFLMD-----DH 213  
DB 315 SAPADSERGPKEPPSGSGPAPRRRRRAQDSSQRRSLPAEBSGSGKDEGGGSSSDH 374  
QY 214 D--QEEPDILKTGLYSKRAAK-----SPDTSDD-----DFWEEGG 246  
DB 375 GGRKKGELPPSSCCRRAGTKSSRSRHSQSSGDESDSDSHRLQKSPQYSEEBE 434  
QY 247 EEDGSGDGMGDSSEFLQDFSETYERY--HTESLQNMKSQELIKYLELEKLSRM-- 301  
DB 435 EEDSGSEH-----SRSRSHSGRRHSHRSRSHSYSS--SDASSQCSYSHQRY 482  
QY 302 -----EDENNRLLSEKRLGDD 319  
DB 483 SDDSYSDYSDRRSHRSKSHSD 505

RESULT 10

A45973  
trichoyalin - human  
C:Species: Homo sapiens (man)  
C>Date: 03-May-1994 #sequence\_revision 01-Mar-1996 #text\_change 22-Jun-1999  
C:Accession: A45973  
R:Lee, S.C.; Kim, I.G.; Marekov, L.N.; O'Keefe, E.J.; Parry, D.A.D.; Steinert, P.M.  
J. Biol. Chem. 268, 12164-12176, 1993  
A:Title: The structure of human trichoyalin. Potential multiple roles as a functional E  
ed (cross-linking) protein.  
A:Reference number: A45973; PMID:93280194; PMID:765034  
A:Accession: A45973  
A:Molecule type: DNA  
A:Residues: 1-1898 <LEB>  
A:Cross-references: GB:109190; NID:q292835; PIDN:AAA65582.1; PID:q292836  
A>Note: authors translated the codon AGC for residue 1714 as Pro  
C:Comment: Trichoyalin is a protein of the medulla of the hair and of the inner root sh  
Covariant modifications to this protein include conversion of arginine to citrulline and  
C:Genetics:  
A:Gene: GDB:THH  
A:Cross-references: GDB:136223; OMIM:190370  
A:Map position: 1921-1921  
C:Superfamily: trichoyalin; calmodulin repeat homology  
C:Keywords: calcium binding; citrulline; EF hand; hair; tandem repeat  
F:49-81/Domains: calmodulin repeat homology <EF2>

Query Match 6.9%; Score 131; DB 1; Length 1898;  
Best local similarity 22.4%; Pred. No. 3.9;  
Matches 79; Conservative 54; Mismatches 146; Indels 74; Gaps 13;

QY 31 ERPPGAER-----VPEEDSRWQRAFPOLGPRGEGSLSEQPPPLQTOACPESS 83  
DB 813 EROLRAEERQORQORFLPEEEKEQR-----GRORRREKLOPLLEEEOLOQRERRAQ 865  
QY 84 CLREGGKQNGDSSAGSGPPEPAVEPTPEAE-----LAQPCHESEASKLGAAPAG 136  
DB 866 QLOEEDDGLQEDERRRQORQDQKRWOLEEERKRRHTLYKPLQLOLRKEQOLQE 925  
QY 137 GBEWQOQOLQKKKRRRPSKKRRHWKPYKLTWEKKKKFDEKQSLRASIRAEK 196

DB 926 EEEELQREER-----KRRQOERYREEOLOQEEEOLEERERKRRRERYRK 979  
QY 197 GQVAPYNTTQFLMDHQ---EEPDILKTGLYSKRAAKSDTSDDFWEEGGSD 253  
DB 980 DKK-----LQKEEOLOGEPEK-----RQEEKRYREBELOE----- 1016  
QY 254 GNGDGSSEFLQDFSETYER-----YHTESLQNMKSQELIKYLELEKLSRMED--EN 305  
DB 1017 -----EQLLREERKRRQEWERYRKDELQOEEOLLEERERKRLQREERYRE 1070  
QY 306 NLRLESKRLGDDARVR---ELEDEL--DLRLAENLOLTEN-ELHROQR 351  
DB 1071 EELQOEEOLLEERETRRQELERYKKEELOLEEBQLEEBEKRROR 1123

RESULT 11

T34418  
hypothetical protein F12F3.3 - Caenorhabditis elegans  
C:Species: Caenorhabditis elegans  
C>Date: 29-Oct-1999 #sequence\_revision 29-Oct-1999 #text\_change 29-Oct-1999  
C:Accession: T34418  
R:Pulion, B.; Wohlmann, P.  
submitted to the EMBL Data Library, July 1998  
A:Description: The sequence of C. elegans cosmid F12F3.  
A:Reference number: 221521  
A:Accession: T34418  
A:Status: preliminary; translated from GB/EMBL/DBJ  
A:Molecule type: DNA  
A:Residues: 1-3488 <FUL>  
A:Cross-references: EMBL:U80022; PIDN:AAC25885.1; GSPDB:GN00023; CESP:F12F3.3  
A:Experimental source: strain Bristol N2; clone F12F3  
C:Genetics:  
A:Gene: CESP:F12F3.3  
A:Map position: 5  
A:introns: 281/3; 332/1; 562/3; 600/3; 1866/3; 1944/3; 3393/1

Query Match 6.9%; Score 131; DB 2; Length 3488;  
Best local similarity 20.5%; Pred. No. 7.3;  
Matches 73; Conservative 63; Mismatches 150; Indels 70; Gaps 10;

QY 8 EYQHQPQTSNCTGAAY--QEELEPERPPGAERVP-----EEDSRWQRAFPOLGPR 59  
DB 1120 ELEKQALIKKAAGADAVKQKELDEKNLEANKKSAAGKLIEBSAAK-----KQ 1171  
QY 60 GPEGSGSLESQPPLOTQACPESSCLRGEGKQNDSDSAGDPPPAVEPTPEAE 119  
DB 1172 TVEEQAKLDAQ--TKATTAERQTLREKDEKSTKSESSEYDERPKKKVLRKTEKSDS 1228  
QY 120 QPCHESEASKLGAAPAGSEEWGQOQOLGKKKRRRPSKKRRHWKPYKLTWEKKKFD 179  
DB 1229 SLSQSESTKYVESAGSESETQVADAARQ-KETBEKQLEAFITAKKSADSKYLE 1287  
QY 180 EKQSL-RASIRAEEMFAKQOPVAPYNTTQFLMDHQEEDPDILKTGLYSKRAAKSDTSD 238  
DB 1288 AESKLKKAAYEA--AKQ-----KEKQOLKLDTPAASKKAAYEKLIEK 1331  
QY 239 DFWEEGGSEBDGSGMGDGSSEFLQDFSETYERYHTESLQNMKSQELIKYLELEKCL 298  
DB 1332 OSHIKKAAEVDA-----VKQKELEKQRLSEBA 1360  
QY 299 SMEENNRLLSEKRLGDDARVRELELDRLAENLOLTPEELHROQERAPL 354  
DB 1361 ATKQADAEKLEBQKKAALALIEIQKEQKLAQEOSRL--EDAKKSAEKQKL 1414

RESULT 12

T30177  
cytoskeleton assembly control protein homolog slaz - yeast (Yarrowia lipolytica)  
C:Species: Yarrowia lipolytica, Candida lipolytica  
C>Date: 22-Oct-1999 #sequence\_revision 22-Oct-1999 #text\_change 22-Oct-1999  
C:Accession: T30177  
R:Gauemann, U.; Schilabel, M.B.; Kurtschko, C.

submitted to the EMBL Data Library, July 1996  
A:Description: SLA2 homologue of Yarrowia lipolytica.

A:Reference number: 220763

A:Accession: T30177

A:Status: preliminary; translated from GB/EMBL/DBJ

A:Molecule type: DNA

A:Residues: 1-1054 <AAU>

A:Cross-references: EMBL:U65409; NID:G3978133; PID:G3978134; PIDN:AAC8184.1

A:Experimental source: strain W29

C:Genetics:

A:Gene: SLA2

A:Map position: III, adjacent to MATA

Query Match 6.8%; Score 129; DB 2; Length 1054;

Best Local Similarity 23.4%; Pred. No. 2.8;

Matches 92; Conservative 45; Mismatches 124; Indels 132; Gaps 19;

QY 31 ERPPGAERPEESRQSAFPGGRPGE-GLBSQPPLOTQACPBSSCLREGE 89  
DB 262 KNPPNLE---EDG-----PGLPQRPKSVATRSVSPAPLPV-ATPTPSI----- 305

QY 90 KQNGDDSSAGDPPEAVEPT---PEAVLAQCHDEBASKGAPAGGEEMGQ-- 143  
DB 306 -----PAAQPLENFWSDALLAQOQYDAEQERLROQAMBEERTIQMQ 349

QY 144 -QOROLGKKHRRSPKSKHMKVYKLTWEKKKPE---KQSLRARIEMFAKQ 198  
DB 350 MQQQQGEWQORQMEAQQA-----QEQLMADQMARHAGRWALERDILALRGQ 400

QY 199 PVAPYNTTQFLMDHDEEDDLTKGLYSKRAAKSDTSDDFMEGEGEGSDGKGD 258  
DB 401 ----YDDQQLMLEQYDRVVALLEALNLQLOOTAHQSAQAKDDILES----- 442

QY 259 GSEFLQDFSEYRYHTESLQNMWSKQELIKELYLEKCL-----SRM 301  
DB 443 ----LQQITWKRQKETTLLKRYSSWRE---EYALTKKLKATQOKAKAEIKAEKEM 495

QY 302 E-----DENNELRLSKRL-GGDDARVLELEL-----DRLAEN 336  
DB 496 ERMHRKNIELADLKERDARAYDLRAKGNKEDVERLRELRMAQDKLADDRSGAD 555

QY 337 LQLL-----TENELRQDERAPLSKFGD 359  
DB 556 LSLLSKGRNELSELENAL-KMKORA-LDERGD 586

# RESULT 13

B45344 probable nuclear antigen - suid herpesvirus 1 (strain Kaplan)

C:Species: suid herpesvirus 1

C>Date: 30-Sep-1993 #sequence\_revision 30-Sep-1993 #text\_change 16-Jul-1999

C:Accession: B45344

R:Vitek, C.; Kozmik, Z.; Paces, V.; Schirm, S.; Schwyzler, M.

Virology 179, 365-377, 1990

A:Title: Pseudorabies virus immediate-early gene overlaps with an oppositely oriented

A:Reference number: A45344; MUID:91021039; PMID:2171211

A:Accession: B45344

A:Status: translation not shown

A:Molecule type: DNA

A:Residues: 1-1733 <VLC>

A:Cross-references: GB:M34651; NID:G334070; PIDN:AAA47471.1; PID:G334072

C:Superfamily: pseudorabies virus 1 nuclear antigen

Query Match 6.7%; Score 128; DB 1; Length 1733;

Best Local Similarity 22.4%; Pred. No. 5.3;

Matches 74; Conservative 27; Mismatches 94; Indels 136; Gaps 16;

QY 8 EYOHQPTSNCT-----GAAVOEE-----LNPER-PPGAEE----- 39  
DB 104 DHQHRPPTTTTIKDPQHPDRLPTKTLQEDBNHLRPTRPDSAKTHNHDPGG 163

QY 40 VPEDSKWQSRAPPOLGRRP-----EGESLESQPPPLQ----- 75

DB 164 GPPSTSHHHNODPPGGPPPPPPPPSTSSSSSHGCPSTPPPPPPPPPPPPPPKXI 223

QY 76 --TOACRESSCLR-----EGEKQNGDDSSAG-----DPPPAVEEPTPA 115

DB 224 SETRASENTAQTQLFHSNENKLFHSMGEGGE-GDGTAGGEGDDDDPPPPPPPPPP 282

QY 116 ELIAPCHDESKGAPAGGEEWQOQROLG--KKHRRPSKKRHKMKVYKLTW 172

DB 283 PLPPPPPPPPPPQ---PPPAGS---ARRRRGGGPPGGRGRRGGKRRR----- 326

QY 173 EEKKKFKDEKQSLRARIEMFAKQPVAPYNTTQFLMDHDEEDDLTKGLYSKRAAK 232  
DB 327 -----AETEA-----AADEEEDGD----- 344

QY 233 SDTSDDDFMEGEGEGSD--GMGEGSE 261  
DB 345 GDEDEDEDRABEGREDGEGPRGAGGAGE 375

# RESULT 14

106310 hypothetical protein Fl1C18.80 - Arabidopsis thaliana

C:Species: Arabidopsis thaliana (mouse-ear cress)

C>Date: 30-Apr-1999 #sequence\_revision 30-Apr-1999 #text\_change 22-Oct-1999

C:Accession: T06310

R:Sevan, M.; Terry, N.; Ardiles, W.; Buyschaert, C.; Dasseville, R.; De Clerck, R.; De

ewes, H.W.; Mayer, K.F.X.; Schueller, C.

A:Submitted to the Protein Sequence Database, April 1999

A:Reference number: Z15589

A:Accession: T06310

A:Molecule type: DNA

A:Residues: 1-852 <BEV>

A:Cross-references: EMBL:AL049607; GSPDB:GN00062; ATSP:Fl1C18.80

A:Experimental source: cultivar Columbia; BAC clone Fl1C18

C:Genetics:

A:Gene: ATSP:Fl1C18.80

A:Map position: 4

A:Introns: 34/3; 102/3; 163/2; 205/3; 269/3; 282/3; 564/3; 602/3; 622/2; 639/3; 667/3; 68

Query Match 6.7%; Score 127.5; DB 2; Length 852;

Best Local Similarity 23.2%; Pred. No. 2.8; Indels 95; Gaps 18;

Matches 88; Conservative 51; Mismatches 145; Indels 95; Gaps 18;

QY 1 MAEPFLSEYOHQPTSNCTGAAVOEELNPERPPGAEEVPEDSRWQSRAPPOLGRRG 60  
DB 439 MAEE-LPEQSVPKTANQKKKSTEEVAPASIAEE-VSEEPNTSE---PQYTKSG 492

QY 61 PGEGLSOPP-PILOTACPBSSCLREGEKQNGDDSSAGDPPEAVEP---TPRA 115  
DB 493 KTVASSSKTKPTVPPEPKSTSTKVAQSEKTVGSDNAQESTKPEKKKGRGKALDE 552

QY 116 ELIAPCHDE--ASLKGAPAGGEEWQ-----QOROLGKKHRRP---SK 159  
DB 553 ESHHTSSGDNKPRPVAASGLASKSKKAEKQYBESNSNTRKRSISLGQKAGESIVSR 612

QY 160 KKRHW---KPYKLTWE---KKKF-----DEKSLRARIEMFAKQVAPYNTTQ 207  
DB 613 IKVWPMDAQAYKGVESYDAKKGLVYDDGDEI-----LYLNQKWSPLDSE 664

QY 208 FLMD---DHDEBDLKT--GLYSKRA---AAKSDTSDDD----- 240  
DB 665 LSQDEEAAQQTQOEEDASTVSGAGSSKAKATPASKSKTSODDDTASKSKEASREE 724

QY 241 --FMEEGEGEDGSDGMGSEFLQDRSEYRYHTESLQNMWSKQELIKELYLEKCL 298  
DB 725 EASSSESESEEBPPTVKGSSRSKSDIS-----SVSKSGK-----SKAS 765

QY 299 SRMEDENRLRLSKRLG 317  
DB 766 SKKKEPSSKATTSKSKSG 784

RESULT 15

A88188  
 protein C18H9.3 [imported] - Caenorhabditis elegans  
 C:Species: Caenorhabditis elegans  
 C:Date: 10-May-2001 #sequence\_revision 10-May-2001 #text\_change 10-May-2001  
 C:Accession: A88188  
 R:Anonymous, The C. elegans Sequencing Consortium  
 S:Science 282, 2012-2018, 1998  
 A:Title: Genome sequence of the nematode C. elegans: a platform for investigating biology  
 A:Reference number: A75000; PMID:99069613; PMID:9851916  
 A:Note: see websites genome.wustl.edu/gsc/C\_elegans/ and www.sanger.ac.uk/Projects/C\_ele  
 A:Note: published errata appeared in Science 283, 35, 1999; Science 283, 2103, 1999; and  
 A:Accession: A88188  
 A:Status: preliminary  
 A:Molecule type: DNA  
 A:Residues: 1-918 <STO>  
 A:Cross-references: GB:chr II: PID:g722383; GSPDB:GN00020; CESP:C18H9.3  
 C:Genetics:  
 A:Note: weakly similar to Trichohyalin  
 A:Gene: C18H9.3  
 A:Map position: 2

Query Match 6.7%; Score 127.5; DB 2; Length 918;  
 Best Local Similarity 21.0%; Pred. No. 3;  
 Matches 87; Conservative 62; Mismatches 150; Indels 115; Gaps 18;

QY 4 PFLSEYQHOPQTSNCTGAAGAAGVDELNPERPPGAEEVPEEDSRMOSRAFPOLG---RP 59  
 DB 307 PQQQQQQQQORSAPYAPSRQSESTDVNLPITPSPDPAWSNNS---MGGGIFYQP 363  
 QY 60 GPRGEGSLSQPP-----PIQTOACP-----ESSCLREGKQON-- 93  
 DB 364 TPQPAPPIVVEBPVQFYMDPTETRRGFPKQDMVWFKAGYFTDSLRYQGENGEYKT 423  
 QY 94 -GDDSSAGDFPPPAVEP-TPEAEILAQCHDEASKLGAPAAAGEEHW---GQ----- 143  
 DB 424 IGLKRLHSGSTPEFYLEDIEPRPILPSIPSATNPPLPAAFGVNMWSSMGQPTDYY 483  
 QY 144 -----QOROLGKKK-----HRRRPSKK-----KRHWKPYKLTWEKKKFKDEKOSL--- 184  
 DB 484 MMQTNFEQQLVARNRLDDHNRRLAEAEAKMAKFOEAMYPQLTMQBQRVREQLLOK 543  
 QY 185 RASRI-RAEMFAKQPVAPYNTTQFLMDHDQEPDLKTGLYSKRAAASDDTSDDFME 243  
 DB 544 RAEIEKREAESKREEA-----RLQKLEQAREIE---ERKAALAEADRKRREI-- 590  
 QY 244 EGGEEDGGSDGMGDSSEFLQDPSETYERYHTESLQNSKQELI-KEYLELEKLSRME 302  
 DB 591 -----EYNR-----MCEKKKNEIIIAKEADRRMRLEATE 620  
 QY 303 DENNRRLRESKRLGGDDARVRELELELDRLRAENLQULTENELHQOERAPLSK 356  
 DB 621 RERRRLAEBS-----RVAEEKIRDRVRAELBAREREERKRAAERERLAR 666

Search completed: February 5, 2004, 13:31:59  
 Job time : 48 secs

THIS PAGE BLANK (USPTO)

THIS PAGE BLANK (USPTO)

THIS PAGE BLANK (USPTO)

GenCore version 5.1.6  
Copyright (c) 1993 - 2004 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: February 5, 2004, 12:05:33 ; Search time 25 Seconds  
(without alignments)

675.303 Million cell updates/sec

Title: US-09-972-758a-2

Sequence: 1 MAEPFLSEYOHQPTNSCTG.....LTENELHROGRAPLSKFGD 359

Scoring table: BLOSUM62  
Gapop 10.0 , Gapext 0.5

Searched: 127863 seqs, 47026705 residues

Total number of hits satisfying chosen parameters: 127863

Minimum DB seq length: 0

Maximum DB seq length: 200000000

Post-processing: Minimum Match 0%  
Maximum Match 100%

Listing first 45 summaries

Database: SwissProt\_41.\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

## SUMMARIES

Result No.	Score	Query Length	ID	Description
1	164.5	8.6	1224 1	DYNA_CHICK
2	139	7.3	756 1	BPS1_BOVIN
3	137	7.2	536 1	GAG_MLVGB
4	134.5	7.0	1181 1	NKX1_RAT
5	134	7.0	568 1	INVO_RAT
6	133	7.0	538 1	GAG_MLVMO
7	133	7.0	1493 1	ERC6_HUMAN
8	131	6.9	1898 1	TRHY_HUMAN
9	130	6.8	466 1	CMGA_RAT
10	129.5	6.8	1816 1	AF6_HUMAN
11	128	6.7	1733 1	VNUA_PRYKA
12	128	6.7	2476 1	ATRX_MOUSE
13	127.5	6.7	4687 1	PLE1_RAT
14	126.5	6.6	1407 1	TRHY_RABIT
15	125.5	6.6	773 1	DGR8_HUMAN
16	125.5	6.6	1549 1	TRHY_SHEEP
17	125	6.5	2004 1	MOZ_HUMAN
18	124.5	6.5	332 1	PA9A_HUMAN
19	124	6.5	449 1	CMGA_BOVIN
20	123.5	6.5	677 1	SGL_MOUSE
21	123.5	6.5	1850 1	BA2A_MOUSE
22	122.5	6.4	4473 1	PLE1_CRIGR
23	122	6.4	723 1	SSRP_PROME
24	121	6.3	538 1	GAG_MSVMO
25	121	6.3	726 1	BRD3_HUMAN
26	121	6.3	1089 1	Y553_HUMAN
27	120	6.3	1280 1	DYNA_RAT
28	120	6.3	536 1	GAG_MLVGB
29	120	6.3	699 1	NP14_HUMAN
30	120	6.3	1781 1	AK12_HUMAN
31	119.5	6.3	556 1	HIR3_HUMAN
32	119.5	6.3	919 1	INCE_HUMAN
33	118.5	6.2	472 1	YWIE_CAEEL

34	118.5	6.2	764 1	UBP1_RAT	P25977	rattus norv
35	117.5	6.2	372 1	TOLA_HAEIN	P44678	haemophilus
36	117.5	6.2	513 1	DMPI_HUMAN	Q13316	homo sapien
37	117.5	6.2	2142 1	BAT2_HUMAN	P48634	homo sapien
38	117.5	6.2	5147 1	PCLO_HUMAN	Q96670	homo sapien
39	117	6.1	568 1	RGS3_MOUSE	Q96004	mus musculus
40	117	6.1	845 1	NFW_RAT	P12839	rattus norv
41	117	6.1	911 1	CAFA_MOUSE	Q96710	mus musculus
42	116.5	6.1	503 1	DMPI_MOUSE	Q55188	mus musculus
43	116.5	6.1	1281 1	DYNA_MOUSE	Q08788	mus musculus
44	116	6.1	927 1	SLA2_CAEEL	Q02328	caenorhabdi
45	116	6.1	1729 1	TABP_HUMAN	Q9C0C2	homo sapien

## ALIGNMENTS

RESULT 1  
ID DYNA\_CHICK STANDARD; PRT; 1224 AA.  
AC P35458;  
DT 01-JUN-1994 (Rel. 29, Created)  
DT 16-OCT-2001 (Rel. 40, Last sequence update)  
DT 28-FEB-2003 (Rel. 41, Last annotation update)  
DE Dynein-1 (150 kDa dynein-associated polypeptide) (DP-150) (DAP-150)  
DE (P150-glyed).  
DN DCTN1.  
OS Gallus gallus (Chicken).  
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
OC Archosauria; Aves; Neognathae; Galliformes; Phasianidae; Phasianinae;  
OC Gallus.  
OX NCBI\_TaxID=9031;  
RN [1]  
RP SEQUENCE FROM N.A.  
RA Cleveland D.W.;  
RL Submitted (OCT-1999) to the EMBL/GenBank/DBJ databases.  
RN [2]  
RP SEQUENCE OF 172-1224 FROM N.A.  
RC TISSUE=Embryonic brain;  
RX MEDLINE=92098576; PubMed=1836789;  
RA Gill S.R., Schroer T.A., Szilak I., Steiner E.R., Sheetz M.P.,  
RT "Dynein, a conserved, ubiquitously expressed component of an  
RT activator of vesicle motility mediated by cytoplasmic dynein";  
RL J. Cell Biol. 115:1639-1650(1991).  
CC - FUNCTION: DYNEIN IS A MAJOR COMPONENT OF ACTIVATOR I, A 20S  
CC POLYPEPTIDE COMPLEX THAT STIMULATES DYNEIN-MEDIATED VESICLE  
CC TRANSPORT.  
CC - SUBCELLULAR LOCATION: Cytoplasmic.  
CC - ALTERNATIVE PRODUCTS:  
CC Event=Alternative splicing; Named isoforms=1;  
CC Comment=At least 3 isoforms are produced;  
CC Name=1;  
CC IsoId=P35458-1; Sequence=Displayed;  
CC - TISSUE SPECIFICITY: Ubiquitously expressed.  
CC - SIMILARITY: BELONGS TO THE DYNEIN 150 KDA SUBUNIT FAMILY.  
CC - SIMILARITY: Contains 1 CAP-Gly domain.  
CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
CC the European Bioinformatics Institute. There are no restrictions on its  
CC use by non-profit institutions as long as its content is in no way  
CC modified and this statement is not removed. Usage by and for commercial  
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
CC -----  
CC EMBL; X62773; CA44617.2; -  
CC PIR; A41642; A41642.  
CC InterPro; IPR000938; CAP-Gly.  
CC Pfam; PF01302; CAP\_GLY; 1.  
CC PROSITE; PS00845; CAP\_GLY\_1; 1.  
CC PROSITE; PS50245; CAP\_GLY\_2; 1.  
CC Motor protein; Microtubules; Dynein; Coiled coil; Cytoskeleton;

KW Alternative splicing.  
 FT DOMAIN 49 91 CAP-GLY.  
 FT DOMAIN 205 540 COILED COIL (POTENTIAL).  
 FT DOMAIN 936 1042 COILED COIL (POTENTIAL).  
 FT DOMAIN 1081 1117 COILED COIL (POTENTIAL).  
 SQ SEQUENCE 1224 AA; 135562 MM; 03B7FE68E7C01D7 CRC64;

Query Match 8.6%; Score 164.5; DB 1; Length 1224;  
 Best Local Similarity 22.6%; Pred. No. 0.045;  
 Matches 83; Conservative 61; Mismatches 152; Indels 71; Gaps 14;

QY 29 NPERPPAEERVRVEDSRMQRAPFOLGGRPGEGESLSQPPLOTQCPSSCLREG 88  
 DB 106 SPETPSAALKVPRGRSRXAKSGKLGAKP---KKTARRRPPTPTTAPS---G 157  
 QY 89 EKONGDSSAGDPF--PPPAVEPTP-BAELLAOPCHDEASKLGAPAGGEEWQOO 145  
 DB 158 TAPSGSASAGSGEMSSSEPTAQTPLVAPVIPSBLTPVAPAMPSPFKKEENLRSOV 217  
 QY 146 ROLGKK---KHRRRPSKKRHWKPYKLTWEE---KKKFEDEKQSLRASRIR-AEMFAK 196  
 DB 218 RDLLEKLETKIKRKNEDKALKELEKYKIQLEQVQEWKSKMQOQADLQRLKEAKKEAK 277  
 QY 197 GQVAPRINTQFLMDHDQ-EEDPLKTGLYSKRAAKSDTSDDDFMEEGE----- 247  
 DB 278 DALEAKERYWEADVADTADALEMATLDKEMAEBAESLQOEV---DSLKEVEYLTMDEI 334  
 QY 248 -----EDGSDGNGDGESEFLQRFSEYRYHTESLONNSKQELIK-----EYL 292  
 DB 335 LKHEIEKSGDA-----ASSYQVKOLEEONNALKALVAPMLSLASKEQEHV 382  
 QY 293 ELEKCLSRMEDENNRRLRESKRLGGDARVELELDRLR-----AENIQLTEN 343  
 DB 383 KLOKQKMKTELESJLQREKL---QEVYKQAEKTVDELKEQVDALGAEMVETLTER 429  
 QY 344 ELHRQOE 350  
 DB 440 NLDLEEK 446

RESULT 2  
 BPS1\_BOVIN STANDARD; PRT; 756 AA.  
 ID BPS1\_BOVIN  
 AC 006002;  
 DT 01-OCT-1996 (Rel. 34, Created)  
 DT 15-DEC-1998 (Rel. 37, Last sequence update)  
 DT 28-FEB-2003 (Rel. 41, Last annotation update)  
 DE Filensin (Beaded filament structural protein 1).  
 GN BPS1.  
 OS Bos taurus (Bovine).  
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 OC Mammalia; Eutheria; Cetartiodactyla; Ruminantia; Pecora; Bovidae;  
 OC Bovidae; Bovinae; Bos.  
 NCBI\_TaxID=9913;  
 RN [1]  
 RP SEQUENCE FROM N.A., AND PARTIAL SEQUENCE.  
 RC TISSUE=Lens;  
 RX MEDLINE=93260017; Pubmed=8491777;  
 RA Gounari F., Mendes A., Quintan R., Hess J.F., Fitzgerald P.G.,  
 RA Quizon C.A., Georgatos S.D.;  
 RT "Bovine filensin possesses primary and secondary structure similarity  
 RT to intermediate filament proteins";  
 RL J. Cell Biol. 121:847-853(1993).  
 RN [2]  
 RP REVISIONS, AND SEQUENCE FROM N.A.  
 RC TISSUE=Lens;  
 RA Hess J.F.;  
 RL Submitted (DEC-1997) to the EMBL/Genbank/DBJ databases.  
 CC -1- SUBUNIT: ASSOCIATES WITH BPS2.  
 CC -1- SUBCELLULAR LOCATION: MEMBRANE- AND CYTOSKELETON-ASSOCIATED.  
 CC -1- TISSUE SPECIFICITY: LENS.  
 CC -1- SIMILARITY: BELONGS TO THE INTERMEDIATE FILAMENT FAMILY.

CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
 CC the European Bioinformatics Institute. There are no restrictions on its  
 CC use, by non-profit institutions as long as its content is in no way  
 CC modified and this statement is not removed. Usage by and for commercial  
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
 CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).

DR EMBL; X72388; CAA51081.1; -.  
 DR InterPro; IPR001664; IF.  
 DR PROSITE; PS00226; IF; FALSE NEG.  
 KM Intermediate filament; Repeat; Membrane; Coiled coil; Cytoskeleton;  
 KM Eye lens protein; Phosphorylation.  
 FT DOMAIN 1  
 FT DOMAIN 39 38  
 FT DOMAIN 319 755  
 FT DOMAIN 39 73  
 FT DOMAIN 74 82  
 FT DOMAIN 83 182  
 FT DOMAIN 183 199  
 FT DOMAIN 200 318  
 FT DOMAIN 531 621  
 FT REPEAT 531 544  
 FT REPEAT 545 551  
 FT REPEAT 552 565  
 FT REPEAT 566 579  
 FT REPEAT 580 593  
 FT REPEAT 594 607  
 FT REPEAT 608 621  
 FT MOD RES 5  
 SQ SEQUENCE 756 AA; 83001 MM; F86A1820BA8B6109 CRC64;

Query Match 7.3%; Score 139; DB 1; Length 756;  
 Best Local Similarity 22.6%; Pred. No. 0.67;  
 Matches 88; Conservative 32; Mismatches 117; Indels 152; Gaps 16;

QY 26 EELNPERPPGAERYVEEDSRMQRAPFOLGGRPGEGESLSQPP-PLQTOA----- 78  
 DB 415 EEAAPTQORGAPEDVP--DGGKISXAFELGIMIKYKGPPEPPADLYTKGRYVWVS 472  
 QY 79 -----CPESCLRGEGKNGDSSAGDPPPAVEPTP-----BAELLAQ 120  
 DB 473 GGGSFVDPGFCVFSYPANAGVYVSKDDSDS-----VPPDSGVPSPOQPEPPLEEGGPPQ 527  
 QY 121 PCHDSEASKLGAPAGG-----EEWG 142  
 DB 528 EKEDGLKESGCGPEKGPEPEKSGDSYKEEGPPEKGGDGVKEBGGPPKGGDGVKEEG 587  
 QY 143 QQQ-RQLGKKRRRRPSKKRHWKPYKLTWEEKKPKDEKQSLRASRIRAEKPAQOVA 201  
 DB 588 PEGKDDGVKKEGPEPEKGEGLK-----EEEPLOEKED-----GQSP 627  
 QY 202 PYNTTQFLMDHDQEPDLKTGLYSKRAAKSDTSDDDFMEG-----GEED-- 249  
 DB 628 PRPA-----DKGDEKNAKELKGLQKQ-----DDQKEGARGPCPMVAPGEGS 672  
 QY 250 -----GSDGMGDDGSEFLQRFSEYRYHTESLONNSKQEL 287  
 DB 673 TRSGGPQVITLGSSEGHARGSGRLARSPRLAIEKVEWESIKFSTESIQTVEETAV 732  
 QY 288 IKEYLEKCLSRMEDENNRRLRESKRLG 316  
 DB 733 IVEIM-IEK-----TKANKKLG 749

RESULT 3  
 GAG\_MLVCB STANDARD; PRT; 536 AA.  
 ID GAG\_MLVCB  
 AC P27460;  
 DT 01-AUG-1992 (Rel. 23, Created)  
 DT 01-AUG-1992 (Rel. 23, Last sequence update)  
 DT 28-FEB-2003 (Rel. 41, Last annotation update)  
 DE GAG polypeptide [contains: Core protein p15; inner coat protein p12;

DE Core shell protein P30; Nucleoprotein P10.  
 GN GAG.  
 OS Cas-Br-E murine leukemia virus.  
 CC Viruses; Retroid viruses; Retroviridae; Gammaretrovirus.  
 OK NCBI\_TaxID=11792;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RX MEDLINE=91227170; PubMed=1840655;  
 RA Perryman S.M., McAtee F.J., Portis J.L.;  
 RT "Complete nucleotide sequence of the neurotropic murine retrovirus  
 CAS-Br-E."  
 RL Nucleic Acids Res. 19:1707-1707(1991).  
 CC -1- P.T.M. SPECIFIC ENZYMAIC CLEAVAGES IN VIVO YIELD MATURE PROTEINS.  
 CC -1- MISCELLANEOUS: THIS PROTEIN IS SYNTHESIZED AS A GAG-POL  
 CC POLYPROTEIN.  
 CC -1- SIMILARITY: Contains 1 CCHC-type zinc finger.  
 CC -----  
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
 CC the European Bioinformatics Institute. There are no restrictions on its  
 CC use by non-profit institutions as long as its content is in no way  
 CC modified and this statement is not removed. Usage by and for commercial  
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
 CC or send an email to [license@sib-sib.ch](mailto:license@sib-sib.ch)).  
 CC -----  
 CC EMBL; X57540; CAA40759.1; -  
 CC InterPro; IPR000840; Gag\_MA.  
 CC InterPro; IPR002079; Gag\_P12.  
 CC InterPro; IPR003036; Gag\_P30.  
 CC InterPro; IPR001878; ZnF\_CCHC.  
 CC Pfam; PF01140; Gag\_MA; 1.  
 CC Pfam; PF01141; Gag\_P12; 1.  
 CC Pfam; PF02093; Gag\_P30; 1.  
 CC Pfam; PF00098; ZF\_CCHC; 1.  
 CC PRINTS; PR00939; C2HCZNFINGER.  
 CC SMART; SM00343; ZnF\_CCHC; 1.  
 CC PROSITE; PS50158; ZF\_CCHC; 1.  
 CC Core protein; Coat protein; Nucleoprotein; Polyprotein; Myristate;  
 CC Zinc-finger.  
 CC CHAIN 2 129 CORE PROTEIN P15.  
 CC CHAIN 130 214 INNER COAT PROTEIN P12.  
 CC CHAIN 215 477 CORE SHELL PROTEIN P30.  
 CC CHAIN 478 536 NUCLEOPROTEIN P10.  
 CC ZN FING 500 517 CCHC-TYPE.  
 CC LIPID 2 2 MYRISTATE (BY SIMILARITY).  
 CC SEQUENCE 536 AA; 60740 MW; 3E0F863A393960AE CRC64;  
 SQ  
 Query Match 7.2%; Score 137; DB 1; Length 536;  
 Best Local Similarity 21.3%; Pred. No. 0.6;  
 Matches 88; Conservative 54; Mismatches 141; Indels 130; Gaps 19;  
 QY 30 PERPPGAEVRVPEDESRMOSRA-----FPQLGGRGPE---GEQS-----LESQPPPD 75  
 DB 105 PLPLPSPAPSLPLPPLPSTSPRSSLYPALPLPSLGAKKPQVLPPDSGGPLDLTDEPPPYR 164  
 QY 76 TQACPSSSCLRESEKONGDSAGDPPPAVEVPTPEALL-----AOPCHDSKASKG 131  
 DB 165 DDEPPPS-----DRRDDGEAPAPGAP-----DSPMASRLRGRELFPVADSTTSQAF 213  
 QY 132 APAAGSEEMGQOQOOLGKKKRRRPSKKRHWK---PYT-----XL----- 170  
 DB 214 PLRSGN-----GQGYVPFSSSDLYNNKNNPSSSEDPGLTALIESVLLTHQP 263  
 QY 171 TWEEKKKF-----DEKQSLRASRIAEVFAKGQVAPNTTQFLMDHDEEPD----- 219  
 DB 264 TMDDCOOLLGTLTGEKORV---LLEARKAVVGEGRPLQPLNEINDAFPLERPMDVY 320  
 QY 220 -----LKTGLYSKRAAKSDTSDDEMBESGEEDGSDGSGDSEFLOR 265  
 DB 321 TQGRNHLVLYRQLLAGLON---AGRSPTNL-----AKYKGLTQGNESPSAFLE 369  
 QY 266 DFESEYRYHTESLQ-----NNSKQELIKVYLELEKCLSRMEDENNRLLESKRLGG-- 317

DB 370 -LKEAVRYRTPYDPOGQETVNSFWQAPDIGRKLRLIED-----LKSRTLGLDV 422  
 QY 318 -----DDARVLELELDRLRAEMVQLITENELHROQERAPL 354  
 DB 423 REAEKIPNKRTPPEERERIKETBEKERRRAEDQEKERDRRRHMSKL 475  
 RESULT 4  
 ID NKX1 RAT STANDARD; PRT; 1181 AA.  
 AC Q9QZM6; Q62932; -  
 DT 28-FEB-2003 (Rel. 41, Created)  
 DT 28-FEB-2003 (Rel. 41, Last sequence update)  
 DT 28-FEB-2003 (Rel. 41, Last annotation update)  
 DE Sodium/potassium/calcium exchanger 1 precursor (Na(+)/K(+)/Ca(2+)-  
 DE exchange protein 1) (Retinal rod Na-Ca+K exchanger).  
 GN SLC24A1 OR NCKX1.  
 OS Rattus norvegicus (Rat).  
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Rattus.  
 OK NCBI\_TaxID=10116;  
 RN [1]  
 RP SEQUENCE FROM N.A., AND ALTERNATIVE SPLICING.  
 RC STRAIN=Sprague-Dawley; TISSUE=Eye;  
 RX MEDLINE=20217335; PubMed=10751314;  
 RA Poon S., Leach S., Li X.-F., Tucker J.E., Schuetkamp P.P.M.,  
 RA Lytton J.;  
 RT "Alternatively spliced isoforms of the rat eye  
 RT sodium/calcium-potassium exchanger NCKX1."  
 RL Am. J. Physiol. 278:C651-C660(2000).  
 RN [2]  
 RP SEQUENCE OF 1067-1155 FROM N.A.  
 RA White K.E., Geseck P.A., Friedman P.A.;  
 RL Submitted (FEB-1996) to the EMBL/Genbank/DBJ databases.  
 CC -1- FUNCTION: Critical component of the visual transduction cascade,  
 CC controlling the calcium concentration of outer segments during  
 CC light and darkness. Light causes a rapid lowering of cytosolic  
 CC free calcium in the outer segment of both retinal rod and cone  
 CC photoreceptors and the light-induced lowering of calcium is caused  
 CC by extrusion via this protein which plays a key role in the  
 CC process of light adaptation. Transports one Ca(2+) and one K(+) in  
 CC exchange for four Na(+).  
 CC -1- SUBCELLULAR LOCATION: Integral membrane protein.  
 CC -1- ALTERNATIVE PRODUCTS:  
 CC Event-Alternative splicing; Named isoforms=4;  
 CC Name=1;  
 CC IsoId=Q9QZM6-1; Sequence=Displayed;  
 CC Name=2;  
 CC IsoId=Q9QZM6-2; Sequence=VSP\_006161;  
 CC Name=3;  
 CC IsoId=Q9QZM6-3; Sequence=VSP\_006162;  
 CC Name=4;  
 CC IsoId=Q9QZM6-4; Sequence=VSP\_006163;  
 CC -1- TISSUE SPECIFICITY: Highly expressed in the eye.  
 CC -1- SIMILARITY: BELONGS TO THE SLC24A FAMILY OF TRANSPORTERS.  
 CC -----  
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
 CC the European Bioinformatics Institute. There are no restrictions on its  
 CC use by non-profit institutions as long as its content is in no way  
 CC modified and this statement is not removed. Usage by and for commercial  
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
 CC or send an email to [license@sib-sib.ch](mailto:license@sib-sib.ch)).  
 CC -----  
 CC EMBL; AF176688; AAD53121.1; -  
 CC EMBL; U49235; AAB37753.1; -  
 CC InterPro; IPR004817; K\_NaCaexchang.  
 CC InterPro; IPR004837; NaCa\_Exmemb.  
 CC Pfam; PF01699; NaCa\_Ex; 2.  
 CC TIGRFAMs; TIGR00937; ZAI904; 1.  
 CC TIGRFAMs; TIGR00367; TIGR00367; 1.  
 KW Vision; Transport; Antipport; Symport; Calcium transport;



```

KW Transmembrane; Glycoprotein; Phosphorylation; Signal; Repeat;
KW Alternative splicing.
FT SIGNAL 1 38 POTENTIAL.
FT CHAIN 39 1181 SODIUM/POTASSIUM/CALCIUM EXCHANGER 1.
FT DOMAIN 39 419 EXTRACELLULAR (POTENTIAL).
FT TRANSMEM 420 440 POTENTIAL.
FT TRANSMEM 441 464 CYTOPLASMIC (POTENTIAL).
FT TRANSMEM 465 485 POTENTIAL.
FT DOMAIN 486 491 EXTRACELLULAR (POTENTIAL).
FT TRANSMEM 492 512 POTENTIAL.
FT DOMAIN 513 519 CYTOPLASMIC (POTENTIAL).
FT TRANSMEM 520 544 POTENTIAL.
FT DOMAIN 545 552 EXTRACELLULAR (POTENTIAL).
FT TRANSMEM 553 569 POTENTIAL.
FT DOMAIN 570 589 CYTOPLASMIC (POTENTIAL).
FT TRANSMEM 990 1010 POTENTIAL.
FT DOMAIN 1011 1017 EXTRACELLULAR (POTENTIAL).
FT TRANSMEM 1018 1038 POTENTIAL.
FT TRANSMEM 1039 1053 CYTOPLASMIC (POTENTIAL).
FT TRANSMEM 1054 1074 POTENTIAL.
FT DOMAIN 1075 1092 EXTRACELLULAR (POTENTIAL).
FT TRANSMEM 1093 1113 POTENTIAL.
FT DOMAIN 1114 1121 CYTOPLASMIC (POTENTIAL).
FT TRANSMEM 1122 1142 POTENTIAL.
FT DOMAIN 1143 1150 EXTRACELLULAR (POTENTIAL).
FT TRANSMEM 1151 1171 POTENTIAL.
FT DOMAIN 1172 1181 CYTOPLASMIC (POTENTIAL).
FT REPEAT 1061 1092 ALPHA-1.
FT REPEAT 1093 905 ALPHA-2.
FT DOMAIN 730 905 14 X APPROXIMATE TANDEM REPEATS.
FT REPEAT 730 741 1.
FT REPEAT 742 754 2.
FT REPEAT 755 766 3.
FT REPEAT 767 778 4.
FT REPEAT 779 791 5.
FT REPEAT 792 804 6.
FT REPEAT 805 817 7.
FT REPEAT 818 830 8.
FT REPEAT 831 843 9.
FT REPEAT 844 856 10.
FT REPEAT 857 869 11.
FT REPEAT 870 881 12.
FT REPEAT 882 893 13.
FT REPEAT 894 905 14.
FT DOMAIN 952 974 POLY-GU.
FT MOD RES 975 625 PHOSPHORYLATION (POTENTIAL).
FT CARBOHYD 271 271 N-LINKED (GLCNAC. .) (POTENTIAL).
FT VARSPLIC 598 710 Missing (in isoform 2).
FT VARSPLIC 616 710 Missing (in isoform 3).
FT VARSPLIC 652 679 Missing (in isoform 4).
FT SEQUENCE 1181 AA; 129980 MW; B063C1C1193696AB CRC64;

Query Match 7.0%; Score 134.5; DB 1; Length 1181;
Best Local Similarity 22.5%; Pred. No. 1.9;
Matches 67; Conservative 44; Mismatches 160; Indels 27; Gaps 7;

```

```

DB 872 EAEVNEHQETBAGKVEHSGESTAGEKDEHSGSEFQADDTVEKDG--EEBAENAND 929
QY 236 TSDDFMEEGDEGSDGMDGSEFLORDPSEFYERYHTESLQNMKQELKEYLE 293
DB 930 QGETAGKGAAGCGGSD--GGDSSEHDEHDEDEDEDEDEDEDEDEDEDEDEDEDE 985

RESULT 5
ID INVO RAT STANDARD; PRT; 568 AA.
AC P48938;
DT 01-FEB-1996 (rel. 33, Created)
DT 01-FEB-1996 (rel. 33, Last sequence update)
DT 28-FEB-2003 (rel. 41, Last annotation update)
DE Involucrin.
GN IVL.
OS Rattus norvegicus (Rat).
OC Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Rattus.
OX NCBI_TaxId=10116;
RN (1)
RP SEQUENCE FROM N.A.
RC STRAIN=Sprague-Dawley;
RX MEDLINE=9410476; PubMed=8277848;
RA Djian P., Phillips M., Basley K., Huang E., Simon M., Rice R.H.,
RA Green H.;
RT "The involucrin genes of the mouse and the rat: study of their shared
RT repeats.";
RL Mol. Biol. Evol. 10:1136-1149(1993).
CC -1- FUNCTION: Part of the insoluble cornified cell envelope (CE) of
CC stratified squamous epithelia.
CC -1- SUBCELLULAR LOCATION: Cytoplasmic. Constituent of the scaffolding
CC of the cornified envelope.
CC -1- Tissue specificity: Keratinocytes of epidermis and other
CC stratified squamous epithelia.
CC -1- PTM: Substrate of transglutaminase. Specific glutamines or lysines
CC are cross-linked to keratins, desmoplakin and to inter involucrin
CC molecules.
CC -1- SIMILARITY: BELONGS TO THE INVOLUCRIN FAMILY.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation-
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@sib-sib.ch).
CC -----
CC EMBL; L28818; AAA1445.1; -.
CC DR PIR; I61106; I61106.
CC DR InterPro; IPR002360; Involucrin.
CC DR PROSITE; PS00795; INVOLUCRIN; 1.
KW Keratinization; Repeat.
SQ SEQUENCE 568 AA; 67021 MW; E2B07740EF779345 CRC64;

Query Match 7.0%; Score 134; DB 1; Length 568;
Best Local Similarity 20.9%; Pred. No. 0.94;
Matches 93; Conservative 66; Mismatches 169; Indels 118; Gaps 18;

```

Db 303 HLGEOQHOOQOQOESCEPELHKGKQOHOQTOSSSELQKGQOQKPHEDPWLDPEKOKLH 362  
 QY 179 -----DEKOSLRASRIRAE-----FAKGPVPAVYNTTQFLMDHDOEEP 218  
 Db 363 DPELHKGKQOHOQOESSELQVEKKQKHEKPEPELHKGKQOELHPEMDTEQKOKSLHEP 422  
 QY 219 DLKTGLYSKRAAKSDDTSDDFMEGSGEEDGSGDGGSEFLQDFSEYTERHTES 278  
 Db 423 ELHLG-----KQOESHEPMDTEQKOKSLYEPELHKGKQOEOQIEYEGYOR--SKS 472  
 QY 279 LQWMSQOELLKEYLELEKLSRMEDENNRL--PLESRGLGDDARVLELELRLRAEN 336  
 Db 473 LNOQLQKQKASRQOELDD--SHLEQEKELLDQRLDQELVANKD-----OLERKQKLEN 524  
 QY 337 L-----QLTENEELHROQERAPL 354  
 Db 525 LTOKEKQIKQDLPSTD--RVQETQPI 548

RESULT 6  
 GAG\_MLVMO STANDARD; PRT; 538 AA.  
 AC P03332;  
 DT 21-JUL-1986 (Rel. 01, Created)  
 DT 01-FEB-1996 (Rel. 33, Last sequence update)  
 DT 15-SEP-2003 (Rel. 42, Last annotation update)  
 DE GAG polyprotein [Contains: Core protein P15; Inner coat protein P12;  
 DE Core shell protein P30; Nucleoprotein P10].  
 GN GAG.  
 OS Moloney murine leukemia virus.  
 OC Viruses; Retroid viruses; Retroviridae; Gammaretrovirus.  
 OC NCBI\_TaxID=11801;  
 RN [1]  
 RP SEQUENCE FROM N.A. (CLONE PMLV-1).  
 RX MEDLINE=82035843; PubMed=6169994;  
 RA Shimmick T.M., Lerner R.A., Sutcliffe J.G.;  
 RT "Nucleotide sequence of Moloney murine leukaemia virus";  
 RL Nature 293:543-548(1981).  
 RN [2]  
 RP SEQUENCE OF 2-31, AND MYRISTOYLATION.  
 RX MEDLINE=83169654; PubMed=6340098;  
 RA Henderson L.E., Kruttsch H.C., Oroszian S.;  
 RT "Myristyl amino-terminal acylation of murine retrovirus proteins: an  
 RT unusual post-translational proteins modification";  
 RL Proc. Natl. Acad. Sci. U.S.A. 80:339-343(1983).  
 RN [3]  
 RP SEQUENCE OF 479-529.  
 RX MEDLINE=81264245; PubMed=6267042;  
 RA Henderson L.E., Copeland T.D., Sowder R.C., Smythers G.W.;  
 RA Oroszian S.;  
 RT "Primary structure of the low molecular weight nucleic acid-binding  
 RT proteins of murine leukemia viruses";  
 RL J. Biol. Chem. 256:8400-8406(1981).  
 CC -1- MISCELLANEOUS: THIS PROTEIN IS SYNTHESIZED AS A GAG-POL  
 CC POLYPROTEIN.  
 CC -1- SIMILARITY: Contains 1 CCHC-type zinc finger.  
 CC  
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
 CC the European Bioinformatics Institute. There are no restrictions on its  
 CC use by non-profit institutions as long as its content is in no way  
 CC modified and this statement is not removed. Usage by and for commercial  
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
 CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
 CC  
 CC EMBL; J02255; AAB59942.1; -  
 CC PIR; A03930; FOMVIM.  
 CC PDB; 1A6B; 02-NOV-99.  
 CC PDB; 1MN8; 14-JAN-03.  
 CC InterPro; IPR000840; Gag\_MA.  
 CC InterPro; IPR002079; Gag\_P12.  
 CC InterPro; IPR003036; Gag\_P10.  
 CC InterPro; IPR001878; Znf\_CCHC.

DR Pfam; PF01140; Gag\_MA; 1.  
 DR Pfam; PF01141; Gag\_P12; 1.  
 DR Pfam; PF02093; Gag\_P30; 1.  
 DR Pfam; PF00098; Zf\_CCHC; 1.  
 DR PRINTS; PR00939; C2HCZFINGER.  
 DR SMART; SM00343; Znf\_CCHC; 1.  
 DR PROSITE; PS50158; ZF\_CCHC; 1.  
 KW Core protein; Coat protein; Nucleoprotein; Polyprotein; Myristate;  
 KW Zinc-finger; 3D-structure.  
 FT CHAIN 2 131 CORE PROTEIN P15.  
 FT CHAIN 132 215 INNER COAT PROTEIN P12.  
 FT CHAIN 216 478 CORE SHELL PROTEIN P30.  
 FT CHAIN 479 534 NUCLEOPROTEIN P10.  
 FT ZN FING 502 519 CCHC-TYPE.  
 FT LIPID 2 MYRISTATE.  
 SQ SEQUENCE 538 AA; 60858 MW; 8A7652439B464495 CRC64;

Query Match 7.0%; Score 133; DB 1; Length 538;  
 Best Local Similarity 20.9%; Pred. No. 1;  
 Matches 89; Conservative 40; Mismatches 142; Indels 154; Gaps 19;

QY 30 PERPPGA-----EERVPEEDSRMOSRAFPOLGRPGPE---GEGS-----LESQPPPL 74  
 Db 107 PPLPSPAPSLPLEPPSTPPRSSLYPALT-PSLGAKPKQVYSDSGPLDLTDEPPV 165  
 QY 75 QTOACPESSCLREGEKQNGDSSAGDPPPAVEPTPEAIL-----AOPCHDSASKL 130  
 Db 166 RDRPRPPSD--RDG-----NGEATPAGAP-----DPSFASRLRGRREPPVADSTTSCA 214  
 QY 131 GAPAGGEEBWGQOQOLGKKRRRRPSKKRHWKYKLTWEKKKKFDEKOSLRASRIR 190  
 Db 215 PPLRAGN-----GQLQYMPFSSSDLYWK-----NNNPSFSDPGKLTALIE 257  
 QY 191 AEMFAKGPVAPYNTTQFLMDHDOEPLDKTG-----LYSKRAAKSD----- 235  
 Db 258 SVLI-----THOPTDDCOQLGLTLTBEKQVLLERKAVRGDGPPTQLPNE 307  
 QY 236 -----TSDDFMEEGGE-----DGSDGMDGDS 261  
 Db 308 VDAAPFLERPDWYTTQAGNHLVHYRQLLAGLQNAGRSPNLAKVKGITQGNESPSA 367  
 QY 262 FLORDSEYTER-----HESLQMSKQELIKE 290  
 Db 368 FLRR-LKEAVRRRTYPPDEPQGETVNSFWQSPADIGRKLERLEDLNKTLGDIVR- 425  
 QY 291 YLELEKLSRMEDENNRLRLRSKRLGDDARV-RELELELRLRAENQLTENEELHROQ 349  
 Db 426 --DAEKTFNRRTPEER-----ERIRRETEKEKERRRTEDEQEKERRRRHR 472  
 QY 350 ERAPL 354  
 Db 473 EMKTL 477

RESULT 7  
 ERCC\_HUMAN STANDARD; PRT; 1493 AA.  
 AC Q03468;  
 DT 01-OCT-1993 (Rel. 27, Created)  
 DT 01-OCT-1993 (Rel. 27, Last sequence update)  
 DT 15-SEP-2003 (Rel. 42, Last annotation update)  
 DE Excision repair protein ERCC-6 (Cockayne syndrome protein CSB).  
 GN ERCC6 OR CSB.  
 OS Homo sapiens (Human).  
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 OC Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.  
 OC NCBI\_TaxID=9606;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RX MEDLINE=93092214; PubMed=1339317;  
 RA Trolejstra C., van Gool A., de Wit J., Vermeulen W., Bootsma D.,  
 RA Hoeijmakers J.H.J.;  
 RT "ERCC6, a member of a subfamily of putative helicases, is involved in

RT Cockayne's syndrome and preferential repair of active genes.";  
 RL Cell 71:939-953(1992).  
 RN [2]  
 RP SEQUENCE FROM N.A.  
 RA MEDLINE=93181229; PubMed=8382798;  
 RT "Structure and expression of the excision repair gene ERCC6, involved  
 RL in the human disorder Cockayne's syndrome group B.";  
 RN Nucleic Acids Res. 21:419-426(1993).  
 (3)  
 RP REVIEW ON VARIANTS CSB.  
 RA MEDLINE=99374920; PubMed=10447254;  
 RT Cleaver J.E., Thompson L.H., Richardson A.S., States J.C.;  
 RA "A summary of mutations in the UV-sensitive disorders: Xeroderma  
 RT pigmentosum, Cockayne syndrome, and trichothiodystrophy.";  
 RL Hum. Mutat. 14:9-22(1999).  
 (4)  
 RP VARIANTS CSB, AND VARIANTS.  
 RA MEDLINE=98107940; PubMed=9443879;  
 RT Mallery D.L., Tangenelli B., Colella S., Steingrimsdottir H.,  
 RA van Gool A.J., Trelostra C., Stefanini M., Lehmann A.R.;  
 RT "Molecular analysis of mutations in the CSB (ERCC6) gene in patients  
 RL with Cockayne syndrome.";  
 RN Am. J. Hum. Genet. 62:77-85(1998).  
 (5)  
 RP DISEASE.  
 RA MEDLINE=20206317; PubMed=10739753;  
 RT Meira L.B., Graham J.M. Jr., Greenberg C.R., Busch D.B.,  
 RA Dougherty A.T.B., Ziffer D.W., Coleman D.M., Savre-Trin I.,  
 RT Friedberg E.C.;  
 RA "Wanted: a kindred with original cerebro-oculo-facio-  
 RT skeletal syndrome has a mutation in the Cockayne syndrome B  
 RL (CSB) gene.";  
 RN Am. J. Hum. Genet. 66:1221-1228(2000).  
 (6)  
 RP DISEASE.  
 RA MEDLINE=20231544; PubMed=10767341;  
 RT Colella S., Nardo T., Botta E., Lehmann A.R., Stefanini M.;  
 RA "Identical mutations in the CSB gene associated with either Cockayne  
 RT syndrome or the DeSanctis-Cacchiarelli variant of xeroderma  
 RL pigmentosum.";  
 RN Hum. Mol. Genet. 9:1171-1175(2000).  
 (7)  
 RP FUNCTION: IS INVOLVED IN THE PREFERENTIAL REPAIR OF ACTIVE GENES.  
 RA PRESUMED DNA OR RNA UNWINDING FUNCTION. CORRECTS THE UV SURVIVAL  
 RT AND RNA SYNTHESIS AFTER UV EXPOSURE OF COCKAYNE'S SYNDROME  
 RL COMPLEMENTATION GROUP B.  
 (8)  
 RP SUBUNIT: INTERACTS WITH THE CSB PROTEIN AND A SUBUNIT OF RNA  
 RA POLYMERASE II TRITH.  
 RT  
 RL  
 (9)  
 RP SUBCELLULAR LOCATION: Nuclear (Probable).  
 RA  
 RT  
 RL  
 (10)  
 RP DISEASE: Defects in ERCC6 are the cause of Cockayne's syndrome  
 RA type B (CSB) [MIM:133540]. CSB is a disease which is characterized  
 RT by dwarfism, precociously senile appearance, pigmentary retinal  
 RL degeneration, optic atrophy, deafness, sensitivity to sunlight,  
 and mental retardation.  
 (11)  
 RP DISEASE: Defects in ERCC6 are a cause of cerebro-oculo-facio-  
 RA skeletal syndrome (COPFS) [MIM:214301], a rapidly progressive  
 RT neurological disorder leading to brain atrophy with calcification,  
 RL cataracts, microcornea, optic atrophy, progressive joint  
 contractures, and growth failure. Inheritance is autosomal  
 recessive.  
 (12)  
 RP DISEASE: Defects in ERCC6 are a cause of De Sanctis-Cacchiarelli  
 RA syndrome (DSC) [MIM:278800]; also known as xeroderma idiocy. DSC  
 RT is an autosomal recessive syndrome consisting of xeroderma  
 RL pigmentosum associated with mental retardation, retarded growth,  
 gonadal hypoplasia, and sometimes neurologic complications.  
 (13)  
 RP SIMILARITY: BELONGS TO THE SNF2/RAD54 HELICASE FAMILY.  
 RA  
 RT  
 RL  
 (14)  
 RP DATABASE: NAMB=Atlas Genet. Cyogenet. Oncol. Haematol.;  
 RA WWW="http://www.infobiogen.fr/services/chronocancer/Genes/CSBID302.html".  
 RT  
 RL  
 (15)  
 RP This SWISS-PROT entry is copyright. It is produced through a collaboration  
 RA between the Swiss Institute of Bioinformatics and the EMBL outstation-  
 RT the European Bioinformatics Institute. There are no restrictions on its  
 RL use by non-profit institutions as long as its content is in no way

CC modified and this statement is not removed. Usage by and for commercial  
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
 CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
 CC -----  
 DR EMBL: I04791; AAA52397.1; -  
 DR PIR: A44224; A44224.  
 DR TRAFAC: T04550; -  
 DR Genew: HGNC:3438; ERCC6.  
 DR MIM: 133540; -  
 DR MIM: 214301; -  
 DR MIM: 278800; -  
 DR GO: GO:0003678; F:DNA helicase activity; TAS.  
 DR GO: GO:0003702; F:RNA polymerase II transcription factor acti. .; TAS.  
 DR GO: GO:0006281; P:DNA repair; TAS.  
 DR GO: GO:0006366; P:transcription from Pol II promoter; TAS.  
 DR InterPro: IPR001410; DEAD.  
 DR InterPro: IPR001650; Helicase\_C.  
 DR InterPro: IPR000350; SNF2\_N.  
 DR Pfam: PF00271; helicase\_C\_1.  
 DR Pfam: PF00176; SNF2\_N; 1.  
 DR SMART: SM00487; DEXDC; 1.  
 DR SMART: SM00490; HELIC; 1.  
 KW Helicase; DNA repair; ATP-binding; DNA-binding; Nuclear protein;  
 KW Transcription regulation; Disease mutation; Polymorphism;  
 KW Cockayne's syndrome; Deafness; Dwarfism.  
 KW ASP/GLU-RICH (ACIDIC).  
 KW GLY-RICH.  
 FT DOMAIN 442 446  
 FT DOMAIN 466 481  
 FT DOMAIN 1038 1055  
 FT NP\_BIND 532 539  
 FT SITE 646 649  
 FT VARIANT 255 255  
 FT VARIANT 399 399  
 FT VARIANT 670 670  
 FT VARIANT 851 851  
 FT VARIANT 957 957  
 FT VARIANT 1042 1042  
 FT VARIANT 1095 1095  
 FT VARIANT 1097 1097  
 FT VARIANT 1213 1213  
 FT VARIANT 1413 1413  
 SQ SEQUENCE 1493 AA; 168415 MW; 285257EAE071AC CRC64;  
 Query Match 7.0%; Score 133; DB 1; Length 1493;  
 Best Local Similarity 23.7%; Pred. No. 2.9;  
 Matches 82; Conservative 42; Mismatches 118; Indels 104; Gaps 17;  
 35 GAEEVVP-----EDSRMOSRAFPQLGGRPGEGESLESQPPLOTQACESSECLREG 88  
 201 GAEEVIEIDHNSLEEDA-----EPP-----SLGSLMMPVQETAAEE--LITG 243  
 89 EKGQGDSSAGDPPEAVEPTP-----EA-----ELAQPCHDSEASX 129  
 244 QM-----TPGTQIPQKQKPRIMNEASGFEXYLADQAKSFEKKQGCNNBARX 297  
 130 LGAPR-----AGEEEWQOQPOLGKKHRRRPSKKRHKRPYYKLWEEKKDEKOSL 184  
 298 APAPVTPAPVQNVKQKPKKARVLSKEER-----LKGIKKLQKRALQFGKVGLPKAR 352  
 185 RA-----SIRAEEMPAKGPVAPYNTQPLMDHDE-----EPDL-----KTG 223  
 353 RPWESDMPEABDSEGE--SEYPTEEEEEDEDEGAADADLSGDTYELKPLPRKG 411

QY 224 LYSKRAAKSDTSDDDFMEEGGE-DGSGDMGDSSEFLQDFSTYRTTESIQNM 282  
DB 412 KRCKVPOVE---IDDDFPSSGEBAEAAVSGGGRKGYRDRGDGDYKORLRW 468  
QY 283 SKOELKEYLEKLSRMEDENNRLRLEKRLGGDARRELE 328  
DB 469 NK-----LRLODKERKRLKE-----DSESDEABFD 494

RESULT 8  
TRHY HUMAN STANDARD; PRT: 1898 AA.  
ID TRHY HUMAN  
AC 007283;  
DT 01-OCT-1994 (Rel. 30, Created)  
DT 01-OCT-1994 (Rel. 30, Last sequence update)  
DT 15-SEP-2003 (Rel. 42, Last annotation update)  
DE Trichohyalin.  
GN THH OR TRHY OR THL.  
OS Homo sapiens (human).  
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
OC Mammalia; Eutheria; Primates; Catarrhini; Homidae; Homo.  
OX NCBI\_TaxID=9606;  
RN [1]  
RP SEQUENCE FROM N.A.  
RX MEDLINE=93280194; PubMed=7685034;  
RA Lee S.-C., Kim I.-G., Marekov L.N., O'Keefe E.J., Parry D.A.D.,  
RA Steiner P.M.;  
RT "The structure of human trichohyalin. Potential multiple roles as a  
RT functional EF-hand-like calcium-binding protein, a cornified cell  
RT envelope precursor, and an intermediate filament-associated (cross-  
RT linking) protein."  
RT J. Biol. Chem. 268:12164-12176(1993).  
RN [2]  
RP SEQUENCE OF 1731-1898 FROM N.A., AND CHARACTERIZATION.  
RX MEDLINE=93315897; PubMed=7686953;  
RA O'Keefe E.J., Hamilton E.H., Lee S.-C., Steiner P.M.;  
RA "Trichohyalin: a structural protein of hair, tongue, nail, and  
RT epidermis."  
RT J. Invest. Dermatol. 101:658-716(1993).  
CC -1- FUNCTION: INTERMEDIATE FILAMENT-ASSOCIATED PROTEIN THAT ASSOCIATES  
CC IN REGULAR ARRAYS WITH KERATIN INTERMEDIATE FILAMENTS (KIF) OF THE  
CC INNER ROOT SHEATH CELLS OF THE HAIR FOLLICLE AND THE GRANULAR  
CC LAYER OF THE EPIDERMIS. IT LATER BECOMES CROSS-LINKED TO KIF BY  
CC ISOPEPTIDE BONDS. IT MAY SERVE AS SCAFFOLD PROTEIN, TOGETHER  
CC WITH INVOLUCRIN, IN THE ORGANIZATION OF THE CELL ENVELOPE OR EVEN  
CC ANCHOR THE CELL ENVELOPE TO THE KIF NETWORK. IT MAY BE INVOLVED IN  
CC ITS OWN CALCIUM-DEPENDENT POSTSYNTHETIC PROCESSING DURING TERMINAL  
CC DIFFERENTIATION.  
CC -1- SUBUNIT: Monomer (Probable).  
CC -1- TISSUE SPECIFICITY: FOUND IN THE HARD KERATINIZING TISSUES SUCH AS  
CC THE INNER ROOT SHEATH (IRS) OF HAIR FOLLICLES AND MEDULLA, AND IN  
CC THE FILIFORM PAPILLAE OF DORSAL TONGUE EPITHELIUM (PROBABLE).  
CC -1- DEVELOPMENTAL STAGE: EXPRESSED DURING LATE DIFFERENTIATION OF  
CC THE EPIDERMIS.  
CC -1- DOMAIN: CONSISTS OF NINE DOMAINS. DOMAIN 1 CONTAINS TWO EF-HAND  
CC CALCIUM-BINDING DOMAINS. DOMAINS 2-4, 6, AND 8 ARE ALMOST  
CC ENTIRELY ALPHA-HELICAL, CONFIGURED AS A SERIES OF PEPTIDE REPEATS  
CC OF VARYING REGULARITY, AND ARE THOUGHT TO FORM A SINGLE-STRAINED  
CC ALPHA-HELICAL ROD STABILIZED BY IONIC INTERACTIONS. DOMAIN 6 IS  
CC THE MOST REGULAR AND MAY BIND KIF DIRECTLY BY IONIC INTERACTIONS.  
CC DOMAINS 5 AND 7 ARE LESS WELL ORGANIZED AND MAY INDUCE FOLDS IN  
CC THE MOLECULE. DOMAIN 9 CONTAINS THE C-TERMINUS, CONSERVED AMONG  
CC DIFFERENT SPECIES.  
CC -1- PTM: SUBSTRATE OF TRANSGUTAMINASE. SOME 200 ARGININES ARE  
CC PROBABLY CONVERTED TO CITRULLINES BY PEPTIDYLARGININE DEIMINASE.  
CC -1- SIMILARITY: IN THE N-TERMINAL SECTION, BELONGS TO THE S-100  
CC FAMILY.  
CC -1- SIMILARITY: Contains 2 EF-hand calcium-binding domains.  
CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
CC between the Swiss Institute of Bioinformatics and the EMBL Outstation-  
CC the European Bioinformatics Institute. There are no restrictions on its  
CC use by non-profit institutions as long as its content is in no way

CC modified and this statement is not removed. Usage by and for commercial  
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
CC -----  
DR EMBL: L09190; AAA65582.1; -  
DR PIR: A45973; A45973.  
DR HSSP: P02633; 4ICB.  
DR Genew: HGNC:11791; THH.  
DR MIM: 190370; -  
DR GO: GO:0005856; C:cytoskeleton; NAS.  
DR GO: GO:0005509; F:calcium ion binding activity; TAS.  
DR InterPro: IPR001751; CapB\_S100.  
DR InterPro: IPR002048; EF-hand.  
DR InterPro: IPR002017; Spectrin.  
DR Pfam: PF00036; efhand.1.  
DR Pfam: PF01023; S\_100; 1.  
DR ProDom: PD003407; CapB\_S100; 1.  
DR PROSITE: PS00018; EF\_HAND.1.  
DR PROSITE: PS00303; S100-CAPB.1.  
KW Keratinization; Calcium-binding; Repeat; Cytullination.  
FT DOMAIN 1 91  
FT CA\_BIND 22 33  
FT CA\_BIND 62 73  
FT DOMAIN 314 390  
FT REPEAT 314 336  
FT REPEAT 327 339  
FT REPEAT 340 351  
FT REPEAT 352 364  
FT REPEAT 365 377  
FT REPEAT 378 390  
FT REPEAT 391 396  
FT REPEAT 397 402  
FT REPEAT 403 408  
FT REPEAT 409 414  
FT REPEAT 415 420  
FT REPEAT 421 426  
FT REPEAT 427 432  
FT REPEAT 433 438  
FT REPEAT 439 444  
FT DOMAIN 444 702  
FT DOMAIN 923 952  
FT REPEAT 923 952  
FT REPEAT 953 982  
FT REPEAT 983 1012  
FT REPEAT 1013 1042  
FT REPEAT 1043 1072  
FT REPEAT 1073 1102  
FT REPEAT 1103 1132  
FT REPEAT 1133 1162  
FT DOMAIN 1250 1849  
FT CONFLICT 1752 1752  
FT CONFLICT 1794 1801  
FT CONFLICT 1857 1857  
FT CONFLICT 1880 1880  
SQ SEQUENCE 1899 AA; 247219 MW; A74B5947FB62E31D CRC64;  
Query Match 6.9%; Score 131; DB 1; Length 1898;  
Best Local Similarity 22.4%; Pred. No. 4.7; Mismatches 146; Indels 74; Gaps 13;  
Matches 79; Conservative 54;  
QY 31 ERPPGAER-----VPEDSRMOSRAPPOLGRRPGEGSLSEGPPLPTOACPRSS 83  
DB 813 EROLRAEERQOREORLPBEBEKOR-----GRORREKELQPLSEBOLQRRERAO 865  
QY 84 CLREGEKQNGDSSAGDPFPPAEVEPTPEAE-----LLAOPCHDEASKLGAIPAAG 136  
DB 866 QIQEEDGQGEORRRORQDRDQKRWQLSEBRKRRTLVAKPALQQLKKEQQLQE 925  
QY 137 GEEWQOQROLQKKHRRPYSKTKRHWKPYKLTWEKKKDEKOSLPASATPAMPAK 196  
DB 926 EEEELQREER-----KRRROEROYREBOLQOEBOULREBERKRROREROYRK 979

QY 197 GQVAPYNTTQFLMDHDQ---EEPDKTGLYSRAAKSDTSDDDFMEEGGEDGSD 253  
 DB 980 DKR-----LQKEEQLLGEPEKX-----RQREKKYREEBELQGE----- 1016  
 QY 254 GMGDSGEFLORDSSERYER-----YHTESLQMSQOELLKYLEKLSRMD--EN 305  
 DB 1017 -----BEQLREERERRRQERQYRKDELOQEEQLLREERERRRQERQYREE 1070  
 QY 306 NRRLRSKRLGDDARV---ELELEL---DLRAENILQLTEN-ELHROER 351  
 DB 1071 BELQEEQLLGERETRRQELERQYRKDELOQEEQLLREERERRRQER 1123

RESULT 9  
 CMGA RAT STANDARD; PRT; 466 AA.  
 ID CMGA RAT  
 AC P10354;  
 DT 01-MAR-1989 (Rel. 10, Created)  
 DT 01-MAR-1989 (Rel. 10, Last sequence update)  
 DT 28-FEB-2003 (Rel. 41, Last annotation update)  
 DE Chromogranin A precursor (Cga) [contains: Pancreastatin; Beta-granin;  
 WE-14].  
 GN CHGA.  
 OS Rattus norvegicus (Rat).  
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Rattus.  
 OX NCBI\_TaxID=10116;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RX MEDLINE=8811232; PubMed=2828116;  
 RA Iicangelo A., Okayama H., Eiden L.E.;  
 RT "Primary structure of rat chromogranin A and distribution of its  
 RT mRNA";  
 RL FEBS Lett. 227:115-121(1988).  
 RN [2]  
 RP SEQUENCE OF 29-466 FROM N.A.  
 RC TISSUE=Pancreas;  
 RX MEDLINE=88312980; PubMed=3044825;  
 RA Hutton J.C., Nielsen E., Kasteren W.;  
 RT "The molecular cloning of the chromogranin A-like precursor of beta-  
 RT granin and pancreastatin from the endocrine pancreas";  
 RL FEBS Lett. 236:269-274(1988).  
 RN [3]  
 RP SEQUENCE OF 19-32.  
 RX MEDLINE=85285598; PubMed=3896848;  
 RA Hutton J.C., Hansen F., Peshavaria M.;  
 RT "beta-granins: 21 kDa co-secreted peptides of the insulin granule  
 RT closely related to adrenal medullary chromogranin A.";  
 RL FEBS Lett. 188:336-340(1985).  
 CC -1- FUNCTION: PANCREASTATIN STRONGLY INHIBITS GLUCOSE INDUCED INSULIN  
 CC RELEASE FROM THE PANCREAS.  
 CC -1- SUBCELLULAR LOCATION: Neuroendocrine and endocrine secretory  
 CC granules.  
 CC -1- PFM: CGA IS O-GLYCOSYLATED.  
 CC -1- MISCELLANEOUS: BINDS CALCIUM WITH A LOW-AFFINITY.  
 CC -1- SIMILARITY: BELONGS TO THE CHROMOGRANIN / SECRETOGRANIN PROTEIN  
 CC FAMILY.  
 CC -----  
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation-  
 CC the European Bioinformatics Institute. There are no restrictions on its  
 CC use by non-profit institutions as long as its content is in no way  
 CC modified and this statement is not removed. Usage by and for commercial  
 CC entities requires a license agreement (See http://www.isb-sib.ch/announce/  
 CC or send an email to license@isb-sib.ch).  
 CC -----  
 CC EMBL; X06832; CAA29988.1; -  
 DR HSSP; P05059; ICFK.  
 DR InterPro; IPR001819; Chromogranin\_AB.  
 DR InterPro; IPR001990; Granin.  
 DR Pfam; Pfold71; Granin.  
 DR PRINTS; PR00659; CHROMOGRANIN.

DR PROSITE; PS00422; GRANINS\_1; 1.  
 DR PROSITE; PS00423; GRANINS\_2; 1.  
 KW Signal; Amidation; Calcium-binding.  
 FT SIGNAL 1 18  
 FT CHAIN 19 466 CHROMOGRANIN A.  
 FT PEPTIDE 19 146 BETA-GRANIN  
 FT PEPTIDE 281 332 PANCREASTATIN (POTENTIAL).  
 FT PEPTIDE 361 374 WE-14.  
 FT DOMAIN 92 110 POLY-GLN.  
 FT DOMAIN 231 249 POLY-GLU.  
 FT MOD\_RES 344 350 POLY-GLU.  
 FT MOD\_RES 332 332 AMIDATION (G-333 PROVIDE AMIDE GROUP)  
 FT DISULFID 35 56 (PROBABLY).  
 FT SEQUENCE 466 AA; 52024 MW; 05D135FFA657C48C CRC64;  
 BY SIMILARITY.

Query Match 6.8%; Score 130; DB 1; Length 466;  
 Best Local Similarity 24.0%; Pred. No. 1.3; Indels 62; Gaps 17;  
 Matches 87; Conservative 55; Mismatches 159;  
 QY 3 EEPFLSEYOHQPOTSNCTGAAYVOEELNPPRGAEERYEEDSRMOSRAFPQLGGR---P 59  
 DB 64 ERYLSILRHQ-----NLKELQDLALGAKERRAQQQQQQQQQQQQQQQQQQQQS 113  
 QY 60 GPEGEGS--LESQPPPLQ---TQACPESSCLREBEGKQNGDSSAGG--DFP-PPAEVE 110  
 DB 114 SEDELESEVFENQSPAKGDAASEAPSKDTVEKEDSDKQQDAFEGTTEGPPRQAFPE 173  
 QY 111 PTPPEALL--AQPCDSEASKGAPAPAGEEEMGQOQRLGKKRRRPS--KKXRHMKP 166  
 DB 174 PROESSMKNQSOPGSDTNNNTQSPSLPSQHGIPQTTEGSE--RGPSAQQAARKAKQ 230  
 QY 167 YKLTWEKKKPKDEKQSLRIRAEFAPKQPVAPYNT-----TOFLMDHDQEEPD 220  
 DB 231 EEKEEEKEEKEEKEEKEKAIAREKAGKEVPTAASSHFVSQYKTKQKDDGQSESGA 290  
 QY 221 KTGYSKRAAKSDTSDDDFMEEGGEDGSDG-----GGDSGEFLORDSSERY 271  
 DB 291 VNG---KTGSAVAVSEKGELEHSQGEDEGEAMAGPQGLFPGKQGE--LERQOEEB 346  
 QY 272 ERYHTESLQMSQOELLKYLEKLSRMDENNRRLRSKRLGDDARVLELELDR 331  
 DB 347 E-----EEERISREW--EDKRWSMQDLAKELTAE--RLRGEEDDPRSKMLSF-R 392

QY 332 LRA 334  
 DB 393 ARA 395

RESULT 10  
 AF6 HUMAN STANDARD; PRT; 1816 AA.  
 ID AF6 HUMAN  
 AC P55196; Q75087; Q75088; Q75089; Q2NU92;  
 DT 01-OCT-1996 (Rel. 34, Created)  
 DT 16-OCT-2001 (Rel. 40, Last sequence update)  
 DT 28-FEB-2003 (Rel. 41, Last annotation update)  
 DE AF-6 protein.  
 GN MULT4 OR AF6.  
 OS Homo sapiens (Human).  
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 OC Mammalia; Eutheria; Primates; Catarrhini; Homidae; Homo.  
 OX NCBI\_TaxID=9606;  
 RN [1]  
 RP SEQUENCE FROM N.A. (ISOFORM 1).  
 RX MEDLINE=94061833; PubMed=8242616;  
 RA Praeud R., Gu Y., Alder H., Nakamura T., Canani O., Saito H.,  
 RA Huebner K., Gale R.P., Nowell P.C., Kuriyama K., Miyazaki Y.,  
 RA Croce C.M., Canani E.;  
 RT "Cloning of the ALL-1 fusion partner, the AF-6 gene, involved in  
 RT acute myeloid leukemias with the t(6;11) chromosome translocation";  
 RL Cancer Res. 53:5624-5628(1993).  
 RN [2]  
 RP SEQUENCE FROM N.A., AND ALTERNATIVE SPLICING.

RC TISSUE=Fetal brain;  
 RX MEDLINE=9834442; PubMed=9679199;  
 RA Saito S., Matsushima M., Shitahama S., Minaguchi T., Kanamori Y.,  
 RA Minami M., Nakamura Y.;  
 RT "Complete genomic structure, DNA polymorphisms, and alternative  
 RT splicing of the human AF-6 gene.";  
 RL DNA Res. 5:115-120(1998).  
 RN (3)  
 RP SEQUENCE OF 337-1816 FROM N.A. (ISOFORM 2).  
 RA Williams S.;  
 RL Submitted (FEB-2000) to the EMBL/GenBank/DBJ databases.  
 CC -1- FUNCTION: MAY ACT AS AN INTRACELLULAR SIGNALING COMPONENT  
 CC -1- CONTROLLED BY RAS SIGNALING PATHWAYS.  
 CC -1- SUBUNIT: BINDS DIRECTLY TO ZO-1 AND OCCLUDIN.  
 CC -1- SUBCELLULAR LOCATION: Cytoplasmic (Probable).  
 CC -1- ALTERNATIVE PRODUCTS:  
 CC Event=Alternative splicing; Named isoforms=3;  
 CC Name=2;  
 CC IsoId=P55196-1; Sequence=Displayed;  
 CC Name=1;  
 CC IsoId=P55196-2; Sequence=VSP\_000217, VSP\_000218;  
 CC Name=3;  
 CC IsoId=P55196-3; Sequence=VSP\_000219, VSP\_000220;  
 CC -1- DISABS: Involved in acute leukemias by a chromosomal  
 CC translocation t(6;11)(q27;q23) that involves MLLT4 and MLL/HRX.  
 CC The result is a rogue activator protein.  
 CC -1- SIMILARITY: Contains 1 dilute domain.  
 CC -1- SIMILARITY: Contains 1 PDZ/DHR domain.  
 CC -1- SIMILARITY: Contains 2 Ras-associating domains.  
 CC -1- DATABASE: NAME=Atlas Genet. Cytogenet. Oncol. Haematol.;  
 CC WWW="http://www.infobiogen.fr/services/chromancer/Genes/AF6.html".  
 CC -----  
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
 CC the European Bioinformatics Institute. There are no restrictions on its  
 CC use by non-profit institutions as long as its content is in no way  
 CC modified and this statement is not removed. Usage by and for commercial  
 CC entities requires a license agreement (See http://www.ebi.ac.uk/announcements/  
 CC or send an email to license@ebi.ac.uk).

FT DOMAIN 991 1077 PDZ.  
 FT FT GLU/LYS-RICH.  
 FT DOMAIN 162 174 POLY-PRO.  
 FT FT POLY-PRO.  
 FT DOMAIN 1349 1356 POLY-PRO.  
 FT FT 1371 1376 ASP/GLU-RICH (ACIDIC).  
 FT DOMAIN 1561 1571 MLT FUSION POINT.  
 FT SITE 26 26 (IN AN ACUTE MYELOID LEUKEMIA PATIENT).  
 FT FT  
 FT VARSPLIC 1588 1611 LQDERRRQQQLERKKEAEEDRA -> VKGGVWLCPSSV  
 FT FT PIIASACFPWG (in isoform 1).  
 FT FT /FTId=VSP\_000217.  
 FT VARSPLIC 1612 1816 Missing (in isoform 1).  
 FT FT /FTId=VSP\_000218.  
 FT VARSPLIC 1666 1743 /FTId=VSP\_000219.  
 FT FT /FTId=VSP\_000220.  
 FT FT G -> V (IN REF. 1).  
 FT FT P -> PGRNRHFAVYNYHYE (IN REF. 3).  
 FT FT D -> DSSHFD (IN REF. 3).  
 FT FT 1031 1031 D -> DV (IN REF. 1).  
 FT FT 1408 1408 R -> P (IN REF. 1).  
 FT FT  
 FT VARSPLIC 1744 1816 /FTId=VSP\_000219.  
 FT FT Missing (in isoform 3).  
 FT FT /FTId=VSP\_000220.  
 FT FT G -> V (IN REF. 1).  
 FT FT P -> PGRNRHFAVYNYHYE (IN REF. 3).  
 FT FT D -> DSSHFD (IN REF. 3).  
 FT FT 1031 1031 D -> DV (IN REF. 1).  
 FT FT 1408 1408 R -> P (IN REF. 1).  
 FT FT  
 SQ SEQUENCE 1816 AA; 205604 MW; EB1FE7F04879CE8F CRC64;  
 Query Match 6.8%; Score 129.5; DB 1; Length 1816;  
 Best Local Similarity 20.3%; Pred. No. 5.5; Mismatches 144; Indels 111; Gaps 14;  
 Matches 78; Conservative 51;  
 QY 14 QTSNCTGAAYVEELN--PERPPGAEEERYBEDSRWQS-RAPPOLGCRPGEGSLESQ 70  
 DB 1291 QSSSIDSTSSQEHNLHNSKSVTPASTLTLSGPGKWPRAIPATPVAVSGPIRTDLP 1350  
 QY 71 PPPLTQACPBSSCLREGEKGNGDSSAGCFPPPAVEYEPPEMLLAQCHDEASKL 130  
 DB 1351 PPPPVHYA-----GPDFGWSMDLP--PP-----SANDI 1380  
 QY 131 GAPAGGEEMWQOQROLKKHRRRPSKKHMKRYUKLTPE-EKKPKDEKQSLRASRI 189  
 DB 1381 GLPSA-----QVAARRRRREHQRWEYERKALBEERKRREKREKRGQMKT 1429  
 QY 190 RAEMPAKGQPVAVYNTQF-----LMDHDQDEP-----DLTKGLYSKR 228  
 DB 1430 QS---LNPAFPSPPLTAQCKKPKPSTLQRPQETVARELPQOQPTIERRDLYITVSK 1486  
 QY 229 AAKSDDTSDDDFMEEGGEEDGSDMGDGSSEPLQRPSEFYERYHTESLONMSKQELI 288  
 DB 1487 ELSSGDSLSPPDW-----KRDAKKELERQOQWHIVDM-----LS 1520  
 QY 289 KEYLELEKLSMDEENRLR-----LESKRLGQDAVARELELE-----LDR 332  
 DB 1521 KEIQELQSKPRSAESDRKLRLKLMLEWQFOKRLQESKQKDEDEEDDDVDTMLMORL 1580  
 QY 333 RAENLQLTENELHROQERAPLSK 356  
 DB 1581 EAFRRARLQDERRRRQQLERWK 1604  
 RESULT 11  
 ID VNUA\_PRIVKA STANDARD; PRT; 1733 AA.  
 AC P33485;  
 DT 01-FEB-1994 (rel. 28, Created)  
 DT 01-FEB-1994 (rel. 28, Last sequence update)  
 DT 01-FEB-1994 (rel. 28, Last annotation update)  
 DE Probable nuclear antigen.  
 OS Pseudorabies virus (strain Kaplan) (PRV).  
 OC Viruses; dsDNA viruses, no RNA stage; Herpesviridae;  
 OC Alphaherpesvirinae; Varicellovirinae;  
 OX NCBI\_Taxid=33703;  
 RN [1]

RP SEQUENCE FROM N.A.  
 RX MEDLINE=91021039; PubMed=2171211;  
 RA Vitek C., Kozmik Z., Paces V., Schim S., Schwyzer M.;  
 RT "Pseudorabies virus immediate-early gene overlaps with an oppositely  
 oriented open reading frame: characterization of their promoter and  
 RT enhancer regions."  
 RL Virology 179:365-377(1990).  
 CC -----  
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
 CC the European Bioinformatics Institute. There are no restrictions on its  
 CC use by non-profit institutions as long as its content is in no way  
 CC modified and this statement is not removed. Usage by and for commercial  
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
 CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
 CC -----  
 DR EMBL; M34651; AAA47471.1; -.  
 DR PIR; B45344; B45344.  
 FT DOMAIN 112 117 POLY-THR.  
 FT DOMAIN 179 173 GLY-RICH.  
 FT DOMAIN 192 196 POLY-SER.  
 FT DOMAIN 271 298 POLY-PRO.  
 FT DOMAIN 304 308 POLY-ARG.  
 FT DOMAIN 883 889 POLY-GLY.  
 FT DOMAIN 1398 1405 POLY-GLY.  
 SQ SEQUENCE 1733 AA; 172166 MW; 0C8CD8BE475B85E2 CRC64;  
 Query Match 6.7%; Score 128; DB 1; Length 1733;  
 Best Local Similarity 22.4%; Pred. No. 6.3;  
 Matches 74; Conservative 27; Mismatches 94; Indels 136; Gaps 16;  
 QY 8 EYOHQPOTSNCT-----GAAYOEE-----INPER-PPGAEER----- 39  
 Db 104 DHGHRPPPTTTTITKDPQRQRPDLPTKLEEDPHLRPRDRPSATHTHHQDPG 163  
 QY 40 VPEEDSRMGARFQGLGRPP-----EGEGSLSQPPPLQ----- 75  
 Db 164 GPSTSHHHNHPGPGGPPRPSTSSSHQGPSTRRPPQPPRRWPPSPQK1 223  
 QY 76 --TQACRESSCLR-----EGEKQNDSSAG-----DFPPAVEPPEA 115  
 Db 224 SETRAGSENTAQTFLSHSEKLFSPHVEGGE- GDRGTAGGEGDRDPPPPPPPP 282  
 QY 116 ELTAQPCDSEASLGAAPAGSEEWQOQOLG---KKGRRRPPSKKHMRYUULTW 172  
 Db 283 PLRPPPPPPPPPP---PPRAGS---ARRRRGGPPRGGRKRRGKRR----- 326  
 QY 173 EEKKKPEDEKSLRASRIEMFAKQGVAPVYNTTQFLMDHDDEPDLKTGLYSKRAAK 232  
 Db 327 -----ABGTREA-----AADAEEBEDGD----- 344  
 QY 233 SDDTSDDDFMEEGEENGSD--GMGGDSE 261  
 Db 345 GDEDEDEDRAEGEGREDGEGPGAGGAGE 375  
 RESULT 12  
 ATRX\_MOUSE  
 ID ATRX\_MOUSE STANDARD; PRT; 2476 AA.  
 AC 061687;  
 DT 15-JUL-1999 (Rel. 38, Created)  
 DT 15-JUL-1999 (Rel. 38, Last sequence update)  
 DT 15-SEP-2003 (Rel. 42, Last annotation update)  
 DE Transcriptional regulator ATRX (X-linked nuclear protein)  
 DE (heterochromatin protein 2) (HPI alpha-interacting protein) (HPI-BP38  
 DE protein).  
 GN ATRX OR XNP OR HPIBP2.  
 OS Mus musculus (Mouse).  
 OC Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;  
 OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.  
 OC NCBI\_TaxID=10090;  
 RP [1]  
 RP SEQUENCE FROM N.A.

RX MEDLINE=98213653; PubMed=9545503;  
 RA Pickette D.J., Taatzen A.O., Higgs D.R., Gibbons R.J.;  
 RT "Comparison of the human and murine ATRX gene identifies highly  
 RT conserved, functionally important domains."  
 RL Mamm. Genome 9:400-403(1998).  
 RN [2]  
 RP SEQUENCE OF 325-1176 FROM N.A.  
 RX MEDLINE=97133299; PubMed=8978696;  
 RA Le Douarin B., Nielsen A.L., Garnier J.-M., Ichinose H.,  
 RA Jeannotin F., Loessen R., Chambon P.;  
 RT "A possible involvement of TIF1 alpha and TIF1 beta in the epigenetic  
 RT control of transcription by nuclear receptors."  
 RL EMBL J. 15:6701-6715(1996).  
 RN [3]  
 RP SUBCELLULAR LOCATION, AND ASSOCIATION WITH PERICENTROMERIC  
 RP HETEROCHROMATIN.  
 RX MEDLINE=20040663; PubMed=10570185;  
 RA McDowell T.L., Gibbons R.J., Sutherland H., O'Rourke D.M.,  
 RA Bickmore W.A., Pombo A., Turley H., Gatter K., Pickette D.J.,  
 RA Buckle V.J., Chapman L., Rhodes D., Higgs D.R.;  
 RT "Localization of a putative transcriptional regulator (ATRX) at  
 RT pericentromeric heterochromatin and the short arms of acrocentric  
 RT chromosomes."  
 RL Proc. Natl. Acad. Sci. U.S.A. 96:13983-13988(1999).  
 CC -1- FUNCTION: COULD BE A GLOBAL TRANSCRIPTIONAL REGULATOR. MODIFIES  
 CC GENE EXPRESSION BY AFFECTING CHROMATIN.  
 CC -1- SUBUNIT: PROBABLY BINDS EZH2. BINDS ANNEXIN V IN A CALCIUM AND  
 CC PHOSPHATIDYLCHOLINE/PHOSPHATIDYLSERINE-DEPENDENT MANNER (By  
 CC similarity).  
 CC -1- SUBCELLULAR LOCATION: NUCLEAR. ASSOCIATED WITH PERICENTROMERIC  
 CC HETEROCHROMATIN DURING INTERPHASE AND MITOSIS, PROBABLY BY  
 CC INTERACTING WITH HPI.  
 CC -1- SIMILARITY: BELONGS TO THE SNF2/RAD54 HELICASE FAMILY.  
 CC -1- SIMILARITY: CONTAINS 1 PHD-TYPE ZINC FINGER.  
 CC -----  
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
 CC the European Bioinformatics Institute. There are no restrictions on its  
 CC use by non-profit institutions as long as its content is in no way  
 CC modified and this statement is not removed. Usage by and for commercial  
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
 CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
 CC -----  
 DR EMBL; AF026032; AAC08741.1; -.  
 DR EMBL; X99643; CAA67962.1; -.  
 DR MGD; MGI:103067; Atrx.  
 DR GO; GO:000228; C:nuclear chromosome, IDA.  
 DR InterPro; IPR001410; DEAD.  
 DR InterPro; IPR001650; Helicase\_C.  
 DR InterPro; IPR000330; SNF2\_N.  
 DR InterPro; IPR001841; Znf.Fing.  
 DR Pfam; PF00271; helicase\_C; 1.  
 DR Pfam; PF00176; SNF2\_N; 1.  
 DR SMART; SM00487; DEXDC; 1.  
 KW DNA repair; Nuclear protein; DNA-binding; Helicase; ATP-binding;  
 KW Zinc-finger.  
 KM ZN\_FING 219 267 PHD-TYPE.  
 FT NP\_BIND 1579 1586 ATP (POTENTIAL).  
 FT SITE 1704 1707 DEGR BOX.  
 FT DOMAIN 319 322 POLY-SER.  
 FT DOMAIN 735 738 POLY-SER.  
 FT DOMAIN 1001 1004 POLY-SER.  
 FT DOMAIN 1130 1135 POLY-GLU.  
 FT DOMAIN 1182 1185 POLY-SER.  
 FT DOMAIN 1238 1245 POLY-ASP.  
 FT DOMAIN 1484 1487 POLY-GLU.  
 FT DOMAIN 1924 1931 POLY-GLU.  
 FT DOMAIN 2205 2208 POLY-LYS.  
 FT DOMAIN 2245 2248 POLY-LYS.  
 FT DOMAIN 2403 2408 POLY-GLN.  
 SQ SEQUENCE 2476 AA; 278601 MW; 90A42B790FC4FFAC CRC64;  
 Query Match 6.7%; Score 128; DB 1; Length 2476;



	Best Local Similarity	19.7%; Pred. No.	9.1;	Mismatches	78; Conservative	52; Mismatches	130; Indels	136; Gaps	13;
Oy	43	EDSMQSRAPFOLGGRPEBEE---	GSLSEOPPIQTACPBSSCLRGEGKQGNDSSA	99					
Dd	1142	EDNNKKQKQRTSAKKKTGNTEKRNSLRATPKRQVDTISSSSDIGDDONASAGEESSD	1201						
Oy	100	GGDFPP-----PAVEP----	PFE-----AEIIA	119					
Dd	1202	EOKIKPTENTLVLPSHTFCOSSGDDEALSKSVATVDDDDDNENRIAKMLLEETKA	1261						
Oy	120	QPCHDSASKTGAPAAAGEEEMGOOQ-----	-POLGKKRRRPSSKK-	161					
Dd	1262	NLSDEDSSDPDEPGCGKKIKGSQEESPADDGELREQLAVOVNSESDSDESCKP	1322						
Oy	162	--RHMKPYKLTWEEKKFDEKQSILRASRIFAEMFAKOVPAYVTTOFLMDHDHQQEBD	219						
Dd	1322	RYRRRLHLHKLTLLSDSGSEGEEK----	PKPKKHKEAKGR	1356					
Oy	220	LKTGLYSRAAALKSDTDDDPMERGGEEDGSDMGDCSFLODRSETERYKHYTESL	279						
Dd	1357	-----NRKRVSSSEDSTDFQESGVSEVSE-----	SEDECRPRTRSKAKELEN	1402					
Oy	280	QNMSKQELIKXYLEIEKCISR-----	MEDENNRLRESKRL	315					
Dd	1403	QRSTRKQKRRRKIRIKQOEISSSENKSHSEDKKGCEBEDDEDEDBEND---DSKSP	1455						
Oy	316	GGDARARELELEDRLRAENLQTLTENELHROOER	351						
Dd	1459	GKGRKKTKIRKI-LKDQKLRTE-----	TQNALMKEBER	1488					
<b>RESULT 13</b>									
ID	PLEI	RAT	STANDARD;	PRT;	4687	AA.			
AC	P304327;	O08879;	O08880;	O08881;					
DT	01-APR-1993	(Rel. 25,	Created)						
DT	01-OCT-1996	(Rel. 34,	Last sequence update)						
DT	28-FEB-2003	(Rel. 41,	Last annotation update)						
De	plectin 1	(PLTN)	(PCN).						
GN	PLEC1.								
OS	Rattus norvegicus (Rat).								
OC	Eukaryota, Metazoa, Chordata, Vertebrata, Euteleostomi;								
XO	Mammalia, Eutheria; Rodentia; Sciurognathi, Muridae; Rattus.								
OX	NCBI_TaxID=10116;								
RN	[1]								
RC	SEQUENCE FROM N.A. (ISOFORM 1).								
RC	TISSUE=Glia; tumor;								
RX	MEDLINE=91268156; PubMed=2050743;								
RA	Wiche G., Becker B., Luber K., Weltzer G., Castanon M.J.,								
RT	Hauptmann R., Stralow C., Stewart M.,								
RT	"Cloning and sequencing of rat plectin indicates a 466-kD polypeptide								
RT	chain with a three-domain structure based on a central alpha-helical								
RT	coiled coil."								
RL	J. Cell Biol. 114:83-99(1991).								
LN	[2]								
RP	REVIEWS.								
RC	TISSUE=Glia; tumor;								
RX	MEDLINE=96210632; PubMed=8633055;								
RA	Liu C.-G., Maercker C., Castanon M.J., Hauptmann R., Wiche G.,								
RT	"Human plectin: organization of the gene, sequence analysis, and								
RT	prosome localization (8q24)."								
RT	Proc. Natl. Acad. Sci. U.S.A. 93:4278-4283(1996).								
LN	[3]								
RP	PARTIAL SEQUENCE FROM N.A. (ISOFORMS 2; 3 AND 4), AND TISSUE								
RP	SPECIFICITY.								
RC	TISSUE=Glia; tumor;								
RX	MEDLINE=97321050; PubMed=9177781;								
RA	Elliotte C.E., Becker B., Oehler S., Castanon M.J., Hauptmann R.,								
RT	Wiche G.,								
RT	"Plectin transcript diversity: identification and tissue distribution								
RT	of variants with distinct first coding exons and rodless isoforms.";								
LN	Genomics 42:115-125(1997).								

```

CC -1- FUNCTION: INTERLINKS INTERMEDIATE FILAMENTS WITH MICROTUBULES AND
CC MICROFILAMENTS AND ANCHORS INTERMEDIATE FILAMENTS TO DESMOZOMES OR
CC HEMIDESMOZOMES. MAY BE INVOLVED NOT ONLY IN THE CROSSLINKING AND
CC STABILIZATION OF CYTOSKELETAL INTERMEDIATE FILAMENTS NETWORK, BUT
CC ALSO IN THE REGULATION OF THEIR DYNAMICS.
CC -1- SUBUNIT: HOMODIMER OR HOMOTETRAMER.
CC -1- ALTERNATIVE PRODUCTS:
CC Event=Alternative splicing; Named Isoforms=4;
CC Name=1;
CC IsoId=P30427-1; Sequence=Displayed;
CC Name=2;
CC IsoId=P30427-2; Sequence=VSP_005050;
CC Name=3;
CC IsoId=P30427-3; Sequence=VSP_005051;
CC Name=4;
CC IsoId=P30427-4; Sequence=VSP_005052;
CC -1- TISSUE SPECIFICITY: WIDELY EXPRESSED WITH HIGHEST EXPRESSION IN
CC SKELETAL MUSCLE AND LOWEST IN THYMUS.
CC -1- DOMAIN: THE N-TERMINUS INTERACTS WITH ACTIN, THE C-TERMINUS WITH
CC VIMENTIN, DESMIN, GFAP, CYTOKERATINS, LAMIN B, WHEREAS BOTH THE N-
CC AND THE C-TERMINUS CAN BIND INTEGRIN BETA-4.
CC -1- PMW: PHOSPHORYLATED BY Cdc2, REGULATES DISSOCIATION FROM
CC INTERMEDIATE FILAMENTS DURING MITOSIS (BY SIMILARITY).
CC -1- SIMILARITY: Contains 1 actin-binding domain.
CC -1- SIMILARITY: Contains 2 calponin-homology (CH) domains.
CC -1- SIMILARITY: Contains 3 plectin repeats.
CC -1- SIMILARITY: Contains 4 spectrin repeats.
CC -1- SIMILARITY: Belongs to the plaklin or cytolinker family.
CC -1- CAUTION: ISOFORM 4 IS A FRAGMENT AT THE N-TERMINUS.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation --
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (see http://www.isb.ch/announce/
CC or send an email to license@isb-ib.ch).
CC -----
CC EMBL; X59601; CAA42169.1; -
CC EMBL; U96274; AAC53209.1; -
CC EMBL; U96275; AAC53210.1; -
CC EMBL; U96276; AAC53211.1; -
CC PIR; A39638; A39638.
CC HSSP; Q01082; 1BKR.
CC InterPro; IPR001589; Actbind.actinin.
CC InterPro; IPR001715; Calponin-like.
CC InterPro; IPR001101; Plectin.repeat.
CC InterPro; IPR005326; S10.plectin.
CC InterPro; IPR002017; Spectrin.
CC Pfam; PF00307; CH; 2.
CC Pfam; PF00681; Plectin; 21.
CC Pfam; PF03501; S10.plectin; 1.
CC ProDom; PD006662; S10.plectin_N; 1.
CC SMART; SM000033; CH; 2.
CC SMART; SM00250; PLEC; 33.
CC SMART; SM00150; SPEC; 6.
CC PROSITE; PS00019; ACTININ_1; FALSE_NEG.
CC PROSITE; PS00020; ACTININ_2; FALSE_NEG.
CC PROSITE; PSS0021; CH; 2.
CC Coiled coil; Repeat; Structural protein; Cytoskeleton; Actin-binding;
CC Phosphorylation; Alternative splicing.
CC KW DOMAIN 1 1473 GLOBULAR 1.
CC FT DOMAIN 1474 2758 CENTRAL FIBROUS ROD DOMAIN.
CC FT DOMAIN 2759 4687 GLOBULAR 2.
CC FT DOMAIN 181 406 ACTIN-BINDING.
CC FT DOMAIN 185 288 CH 1.
CC FT DOMAIN 301 403 CH 2.
CC FT DOMAIN 648 722 SPECTRIN 1.
CC FT REPEAT 743 827 SPECTRIN 2.
CC FT REPEAT 840 933 SPECTRIN 3.
CC FT REPEAT 1318 1418 SPECTRIN 4.
CC FT REPEAT 1472 1692 COILED COIL (POTENTIAL).
CC FT DOMAIN 1724 2760 COILED COIL (POTENTIAL).

```



entities requires a license agreement (See <http://www.isb-sib.ch/announce/> or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).

EMBL: Z19092; CAA79519.1; -  
DR PIR: S28589; S28589.  
DR HSSP: P02633; 41CB.  
DR InterPro: IPR001751; CAPB S100.  
DR InterPro: IPR002048; EF-hand.  
DR Pfam: PF00036; ehand; 1.  
DR Pfam: PF01023; S100; 1.  
DR ProDom: PD003407; CAPB S100; 1.  
DR PROSITE: PS00018; EF\_HAND; 1.  
DR PROSITE: PS00303; S100-CAPB; 1.  
KW Keratinization; Calcium-binding; Repeat; Citrullination.  
FT DOMAIN 1 91 S-100 LIKE.  
FT CA\_BIND 22 33 EF-HAND 1 (LOW AFFINITY) (POTENTIAL).  
FT CA\_BIND 62 73 EF-HAND 2 (HIGH AFFINITY) (POTENTIAL).  
SQ SEQUENCE 1407 AA; 183781 MW; AET1DZAI59F12B7F CRC64;

Query Match 6.6%; Score 126.5; DB 1; Length 1407;  
Best Local Similarity 23.7%; Pred. No. 6.1;  
Matches 83; Conservative 48; Mismatches 124; Indels 95; Gaps 17;

QY 25 QEE-----INPERPPGAEEVPEEDSRWGRAPFQUGRPGEGSLSSQPPLOTQAC 79  
DB 349 QEERRRQLAEEVREQAREGESLITRRWQRLSEAGAR-----QSKVSRPRRQEGS- 402  
QY 80 PESSCIREGKQNGDDSSAGDPFPFAVEFPPEALLAQPCHEASLGAAPAGGE 139  
DB 403 -----RQDERRQDER-----EELEEQ-----ARRQ 427  
QY 140 EW-GGQQRQGLKKRRRPFKKRHWKPYKLTWEKKKDEKQSLRARIEMPAKQ 198  
DB 428 QMAESESERRRRLSARFRLRQLRAER--QEQQRREERERRE-REL----- 479  
QY 199 PVAAYNTQFLMDHQEEDPLTGLYSKRAAKSDTSDPDMEEGEGEGSDGNGD 258  
DB 480 -----QFL-----EEBQL-----QRRERAQQLQEDSDSFOEDRRRRRQEQRPQ 520  
QY 259 GSEFLORDPSEYTERVHT-----ESLQ-NMKQKLLIYELKCLSRMD 303  
DB 521 TWRWQLE--EAQRRTTLVAKQEQEQLEEBELQREKKRQERRRERREK-LQREED 577  
QY 304 ENNRRLRESKRLGDDARVLELELDRLRAENIQLTENEHL--RQER 351  
DB 578 EKRR-RQERQRYRELELQEQQLDRKLREBQLQEREBRLRQER 626

RESULT 15  
DGR8 HUMAN STANDARD; PRT: 773 AA.  
AC Q8WYQ5; Q96G39; Q96GP8; Q9H6L8; Q9H6T7; Q9NRW2;  
DT 28-FEB-2003 (Rel. 41, Created)  
DT 28-FEB-2003 (Rel. 41, Last sequence update)  
DT 15-SEP-2003 (Rel. 42, Last annotation update)  
DE DGR8 protein (Digeorge syndrome critical region 8).  
GN DGR8 OR DGRK6.  
OS Homo sapiens (Human).  
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
OC Mammalia; Eutheria; Primates; Catarrhini; Hominiidae; Homo.  
OX NCBI\_TaxId=9606;  
RN [1]  
RP SEQUENCE FROM N.A. (ISOFORM 1).  
RA Shimizu N., Minosima S., Kawasaki K., Sasaki T., Shiohama A.;  
RT "Homo sapiens DGRK6 on 22q11.2."  
RL Submitted (NOV-2000) to the EMBL/GenBank/DBJ databases.  
RN [2]  
RP SEQUENCE FROM N.A. (ISOFORMS 1 AND 2).  
RC TISSUE=Brain; and Muscle.  
RX MEDLINE=22388257; PubMed=2477932;  
RA Strausberg R.L., Feingold E.A., Grouse L.H., Derge J.G.,  
RA Klausner R.D., Collins F.S., Wagner L., Shenmen C.M., Schuler G.D.,  
RA Altschul S.F., Zeeberg B., Buetow K.H., Schaefer C.F., Bhat N.K.,

RA Hopkins R.F., Jordan H., Moore T., Max S.I., Wang J., Hsieh F.,  
RA Diatchenko L., Marusina K., Farmer A.A., Rubin G.W., Hong L.,  
RA Stapleton M., Soares M.B., Bonaldo M.F., Casavant T.L., Scheetz T.E.,  
RA Brownstein M.J., Uediri T.B., Toshiyuki S., Carninci P., Prange C.,  
RA Raha S.S., Loggellano N.A., Peters G.J., Abramson R.D., Mollath S.J.,  
RA Bosak S.A., McEwan P.J., McKernan K.J., Malek J.A., Gunaratne P.H.,  
RA Richardson S., Morley K.C., Hale S., Garcia A.M., Gay L.J., Hulyk S.W.,  
RA Villalon D.K., Muzny D.M., Sodergren E.J., Lu X., Gibbs R.A.,  
RA Fahey J., Helton E., Kettelman M., Madan A., Rodrigues S., Sanchez A.,  
RA Whiting M., Madan A., Young A.C., Shevchenko Y., Bouffard G.G.,  
RA Blakesley R.W., Touchman J.W., Green E.D., Dickson M.C.,  
RA Rodriguez A.C., Grimwood J., Schmutz J., Myers R.M.,  
RA Buterfield Y.S.N., Krzywicki M.T., Skalska U., Smalhus D.E.,  
RA Scherch A., Schein J.E., Jones S.J.M., Marra M.A.;  
RT "Generation and initial analysis of more than 15,000 full-length  
RT human and mouse cDNA sequences."  
RL Proc. Natl. Acad. Sci. U.S.A. 99:16899-16903(2002).  
RN [3]  
RP SEQUENCE OF 1-247 FROM N.A.  
RC TISSUE=Heart;  
RA Gong L., Milius S., Jen J., Yeh E.T.H.;  
RT "Isolation and characterization of a novel human gene deleted in  
RT Digeorge syndrome." to the EMBL/GenBank/DBJ databases.  
RL Submitted (JUL-1999) to the EMBL/GenBank/DBJ databases.  
RN [4]  
RP SEQUENCE OF 204-773 FROM N.A. (ISOFORM 1).  
RC TISSUE=Hepatosoma;  
RA Kawabata A., Hikiji T., Kobatake N., Inagaki H., Ikema Y., Okamoto S.,  
RA Okitani R., Ota T., Suzuki Y., Obayashi M., Nishi T., Shibahara T.,  
RA Tanaka T., Nakamura Y., Isogai T., Sugano S.;  
RT "NEO human cDNA sequencing project."  
RL Submitted (AUG-2000) to the EMBL/GenBank/DBJ databases.  
CC -1- ALTERNATIVE PRODUCTS:  
CC Name=1;  
CC IsoId=Q8WYQ5-1; Sequence=Displayed;  
CC Name=2;  
CC IsoId=Q8WYQ5-2; Sequence=VSP\_003847, VSP\_003848;  
CC Note=No experimental confirmation available;  
CC -1- DISBASE: May play a part in the etiology of the  
CC velocardiocfacial/Digeorge syndrome (VCFs/DGS), a developmental  
CC disorder characterized by structural and functional palate  
CC anomalies, conotruncal cardiac malformations, immunodeficiency,  
CC hypocalcemia, and typical facial anomalies. Most cases result from  
CC a deletion of chromosome 22q11.2 (the Digeorge syndrome chromosome  
CC region, or DGR8).  
CC -1- SIMILARITY: Contains 2 DBM (double-stranded RNA-binding) domains.  
CC -1- SIMILARITY: Contains 1 WW domain.  
CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
CC the European Bioinformatics Institute. There are no restrictions on its  
CC use by non-profit institutions as long as its content is in no way  
CC modified and this statement is not removed. Usage by and for commercial  
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).

KW RNA-binding; Repeat; Alternative splicing.

FT DOMAIN 301 334 MW.  
 FT DOMAIN 511 578 DREM 1.  
 FT DOMAIN 620 685 DREM 2.  
 FT VARSPLIC 303 304 LP -> VL (in isoform 2).  
 FT VARSPLIC 305 773 /FTId=VSP\_003847.  
 FT VARSPLIC 305 773 Missing (in isoform 2).  
 FT CONFLICT 241 247 /FTId=VSP\_003848.  
 FT CONFLICT 255 256 DDPNDV -> VCWQPL (IN REF. 3).  
 FT CONFLICT 274 274 LC -> TR (IN REF. 2; AAH09323).  
 FT CONFLICT 343 343 P -> L (IN REF. 4).  
 FT CONFLICT 706 706 H -> Y (IN REF. 4).  
 FT CONFLICT 706 706 V -> A (IN REF. 4; BAB15165).  
 SQ SEQUENCE 773 AA; 86045 MW; 72D962BBE32890EC CRC64;

Query Match 6.6%; Score 125.5; DB 1; Length 773;

Best Local Similarity 23.1%; Pred. No. 3.7; Indels 97; Gaps 15;

Matches 74; Conservative 41; Mismatches 108; Indels 97; Gaps 15;

QY 58 RPPPEGGSLSSQPPPLQTOACPESSCLREGEKGQNGDSSAGGDFPP-----PAEVE- 110  
 Db 331 RPYFLGTGSIKIDDPPLSSIPCLHYKKMKDNEBEGSSDLTPSGDVSPVKPLSRSALEF 390  
 QY 111 PTPEAEIL-AQPCDSEASKLGAPAGGEEBEGGQQRQLGKKRRRPSKKK---RHWK 165  
 Db 391 PLDEPDSMGADPGPPDEKPLGAAAFG-----ALGQVAKAVEVCKDESVDLEFR 441  
 QY 166 PYYK-----LTWEKKKKFDEKQSLRASRIAMFAKGVAPRYNTTQFLMD 212  
 Db 442 SYLEKRFDEQVTVKKRTVAERQFNRW---KRQAE---SERPIIPAN--OKLITL 492  
 QY 213 HDQEEPDLKTGLYSKRAAKSDTDDFME-----EGGE- 250  
 Db 493 SVQDAFTKKEFVINPN--GKSEVCILHEVYQVILKVPVYNPFECENPSEPFGASVTIDG 550  
 QY 251 GSDGMGGDGS-----EFLQRP-----SETYERYHTESLQN----- 281  
 Db 551 VTYGSGTASSSKLAKNKAARATLEILIPDKQTSBEKPKDSELEFYNHISTEDSRVYE 610  
 QY 282 -MSKQELIKKEYLEKCLSR 300  
 Db 611 LTKAGILSPYQILHECLKR 630

Search completed: February 5, 2004, 13:29:01  
 Job time : 26 secs

GenCore version 5.1.6  
Copyright (c) 1993 - 2004 CompuGen Ltd.

OM protein - nucleic search, using frame\_plus\_p2n model

Run on: February 5, 2004, 15:09:51 / Search time 488 Seconds  
(without alignments)  
2709.886 Million cell updates/sec

Title: US-09-972-758a-2  
Perfect score: 1910  
Sequence: 1 MAEPFLSEYQHQPOTSNCTG.....LTENELHROGRAPLSTKFGD 359

Scoring table:  
BLOSUM62  
Xgapop 10.0, Xgapext 0.5  
Ygapop 10.0, Ygapext 0.5  
Fgapop 6.0, Fgapext 7.0  
Delop 6.0, Delext 7.0

Searched: 2449703 seqs, 1841816367 residues  
Total number of hits satisfying chosen parameters: 4899406

Minimum DB seq length: 0  
Maximum DB seq length: 200000000  
Post-processing: Minimum Match 0%  
Maximum Match 100%  
Listing first 45 summaries

Command line parameters:  
-MODE=frame\_plus\_p2n.model -DEV=xlh  
-O=/cgn2\_1/USPTO/spool/US09972758/runat\_05022004\_095005\_1309/app\_query.fasta\_1.519  
-DB=Published Applications NA -QPM=fastap -SUFFIX=rnpb -MINMATCH=0.1  
-LOOCL=0 -LOOPEXT=0 -UNITS=bits -START=1 -END=-1 -MATRIX=blosum62  
-TRANS=human40.cdi -LIST=45 -DOCLIGN=200 -THR SCORE=pct -THR MAX=100  
-THR MIN=0 -ALIGN=15 -MODE=LOCAL -OUTFMT=ptc -NORM=ext -HEADSIZE=500 -MINLEN=0  
-MAXLEN=200000000 -USER=US09972758\_@CGN\_1\_1\_391\_@runat\_05022004\_095005\_1309  
-NCPU=6 -ICPU=3 -NO\_MMAP -LARGEQUERY -NEG SCORES=0 -WAIT -DSBLOCK=100  
-LONGLOG -DEV TIMEOUT=120 -WARN TIMEOUT=30 -THREADS=1 -XGAPOP=10 -XGAPEXT=0.5  
-FGAPOP=6 -FGAPEXT=7 -YGAPOP=10 -YGAPEXT=0.5 -DELOP=6 -DELEXT=7

Database : Published Applications NA.\*  
1: /cgn2\_6/ptodata/2/pubpna/US07\_PUBCOMB.seq.\*  
2: /cgn2\_6/ptodata/2/pubpna/US06\_PCT\_NEW\_PUB.seq.\*  
3: /cgn2\_6/ptodata/2/pubpna/US06\_NEW\_PUB.seq.\*  
4: /cgn2\_6/ptodata/2/pubpna/US06\_PUBCOMB.seq.\*  
5: /cgn2\_6/ptodata/2/pubpna/US07\_NEW\_PUB.seq.\*  
6: /cgn2\_6/ptodata/2/pubpna/US07\_PUBCOMB.seq.\*  
7: /cgn2\_6/ptodata/2/pubpna/US08\_NEW\_PUB.seq.\*  
8: /cgn2\_6/ptodata/2/pubpna/US08\_PUBCOMB.seq.\*  
9: /cgn2\_6/ptodata/2/pubpna/US09\_PUBCOMB.seq.\*  
10: /cgn2\_6/ptodata/2/pubpna/US09\_PUBCOMB.seq.\*  
11: /cgn2\_6/ptodata/2/pubpna/US09\_PUBCOMB.seq.\*  
12: /cgn2\_6/ptodata/2/pubpna/US09\_NEW\_PUB.seq.\*  
13: /cgn2\_6/ptodata/2/pubpna/US10\_NEW\_PUB.seq.\*  
14: /cgn2\_6/ptodata/2/pubpna/US10\_PUBCOMB.seq.\*  
15: /cgn2\_6/ptodata/2/pubpna/US10\_PUBCOMB.seq.\*  
16: /cgn2\_6/ptodata/2/pubpna/US10\_NEW\_PUB.seq.\*  
17: /cgn2\_6/ptodata/2/pubpna/US10\_PUBCOMB.seq.\*  
18: /cgn2\_6/ptodata/2/pubpna/US10\_NEW\_PUB.seq.\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

## SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
------------	-------	-------------	--------	----	-------------

1	1910	100.0	2199	9	US-09-745-763-33	Sequence 33, Appl
2	1910	100.0	3624	15	US-10-205-823-171	Sequence 171, App
3	1902	99.6	1080	10	US-09-972-758a-2	Sequence 1, Appl
4	733	38.4	1083	12	US-10-264-237-629	Sequence 629, App
5	691	36.2	414	10	US-09-983-655-4895	Sequence 4895, App
6	629	32.9	461	10	US-09-954-531-813	Sequence 813, App
7	619	32.4	495	11	US-09-918-995-403	Sequence 403, App
8	556.5	29.1	1330	13	US-10-074-799-1223	Sequence 1223, App
9	545	28.5	2048	12	US-10-708-260A-1496	Sequence 1496, App
10	184	9.6	549	11	US-09-991-936-1712	Sequence 1712, App
11	162	8.5	6354	15	US-10-084-817-158	Sequence 158, App
12	162	8.5	7396	12	US-10-028-248A-35	Sequence 35, Appl
13	162	8.5	7396	12	US-10-107-782-85	Sequence 35, Appl
14	159.5	8.4	4041	12	US-10-264-049-543	Sequence 543, App
15	155	8.1	5002	12	US-10-085-117-121	Sequence 321, App
16	155	8.1	5002	12	US-10-085-117-120	Sequence 320, App
17	152	8.0	2421	10	US-09-974-300-7891	Sequence 7891, App
18	147.5	7.7	3773	9	US-09-925-302-47	Sequence 47, Appl
19	147.5	7.7	4913	12	US-10-159-563-34	Sequence 34, Appl
20	147.5	7.7	4913	13	US-10-133-937-34	Sequence 34, Appl
21	147	7.7	7453	13	US-10-117-722-248	Sequence 248, App
22	147	7.7	7453	15	US-10-037-270-248	Sequence 248, App
23	147	7.7	7501	13	US-10-117-722-249	Sequence 249, App
24	147	7.7	7501	15	US-10-037-270-249	Sequence 249, App
25	146.5	7.7	2479	12	US-10-108-260A-1966	Sequence 1966, App
26	145	7.6	4301	13	US-09-917-800A-1666	Sequence 1666, App
27	145	7.6	4301	13	US-10-205-194-126	Sequence 126, App
28	144	7.5	2090	13	US-10-094-749-1350	Sequence 1350, App
29	144	7.5	2090	13	US-09-825-489-12	Sequence 12, Appl
30	142.5	7.5	4714	12	US-10-108-260A-1885	Sequence 1885, App
31	141	7.4	2907	12	US-10-085-117-124	Sequence 124, App
32	141	7.4	2962	11	US-09-291-417-106	Sequence 106, App
33	141	7.4	4221	13	US-10-085-117-123	Sequence 123, App
34	141	7.4	4221	13	US-10-120-988-120	Sequence 120, App
35	140	7.3	5973	13	US-10-094-466-53	Sequence 63, Appl
36	139	7.3	1611	12	US-10-136-728-37	Sequence 37, Appl
37	139	7.3	1691	13	US-10-136-728-35	Sequence 35, Appl
38	139	7.3	6537	15	US-10-175-523-186	Sequence 186, App
39	138.5	7.3	5835	11	US-09-927-597-1	Sequence 1, Appl
40	138.5	7.3	5937	11	US-09-927-597-3	Sequence 3, Appl
41	138.5	7.3	6861	13	US-10-341-434-102	Sequence 102, App
42	138.5	7.3	6861	15	US-10-171-311-161	Sequence 161, App
43	138.5	7.2	6900	15	US-10-171-311-163	Sequence 163, App
44	138	7.2	1960	10	US-09-834-975-833	Sequence 833, App
45	138	7.2	2093	12	US-10-104-047-1666	Sequence 1666, App

## ALIGNMENTS

RESULT 1  
US-09-745-763-33  
Sequence 33, Application US/09745763  
Patent No. US20020065394A1  
GENERAL INFORMATION:

APPLICANT: Jacobs, Kenneth

McCoy, John M.

Lavallie, Edward R.

Collins-Racie, Lisa A.

Evans, Cheryl

Mezberg, David

Treacy, Maurice

Spaulding, Vikki

TITLE OF INVENTION: SECRETED PROTEINS AND POLYNUCLEOTIDES

ENCODING THEM

NUMBER OF SEQUENCES: 219

CORRESPONDENCE ADDRESS:

ADDRESSEE: Genetics Institute, Inc.

SNREET: 87 Cambridgepark Drive

CITY: Cambridge

STATE: MA

COUNTRY: U.S.A.

ZIP: 02140

COMPUTER READABLE FORM:

```

MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patent Release #1.0, Version #1.30
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/09/745,763
FILING DATE: 18-Jun-2000
CLASSIFICATION: <Unknown>
ATTORNEY/AGENT INFORMATION:
NAME: Sprunger, Suzanne A.
REGISTRATION NUMBER: 41,323
TELEPHONE: (617) 498-8284
TELEFAX: (617) 876-5851
INFORMATION FOR SEQ ID NO: 33:
SEQUENCE CHARACTERISTICS:
LENGTH: 2199 base pairs
TYPE: nucleic acid
STRANDEDNESS: double
TOPOLOGY: linear
MOLECULE TYPE: cDNA
SEQUENCE DESCRIPTION: SEQ ID NO: <Unknown>
US-09-745-763-33

Alignment Scores:
Pred. No.: 1,49e-164 Length: 2199
Score: 1910.00 Matches: 359
Percent Similarity: 100.00% Conservative: 0
Beet Local Similarity: 100.00% Mismatches: 0
Query Match: 100.00% Indels: 0
DB: Gaps: 0

US-09-972-758a-2 (1-359) x US-09-745-763-33 (1-2199)

QY 1 MetAlaGluProPheLeuSerGluTyrGlnHisGlnProGlnThrSerAsnCysThrGly 20
DB 707 ATGGCGGACCCATTCTTGTGAGATATCAACACACCTCAACCTGTAACAGGT 766
QY 21 AlaAlaAlaValGlnGlnGlnLeuAsnProGluArgProGluValAlaGluGlnArgVal 40
DB 767 GCTGCTGCTGCTCCAGAAAGAGCTGAACCTTGAGCGCCCTCCAGGCGCGAGAGCGGGT 826
QY 41 ProGluGlnAspSerArgTTrpGlnSerArgAlaPheProGlnLeuGlyGlyArgProGly 60
DB 827 CCGAGAGAGACAGTGTGAGTGTGCAATGAGAGCGCTCCCAAGTTGGGTGCGCTCGGGG 886
QY 61 ProGluGlnGlyGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGln 80
DB 887 CCGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 946
QY 81 GluSerSerCysLeuArgGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGln 100
DB 947 GAATCTAGCTGCTGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1006
QY 101 GlnAspPheProProProAlaGlnValGlnProThrProGlnValAlaGlnLeuLeuAlaGln 120
DB 1007 GCGGACTTCCCGCGCGCGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1066
QY 121 ProCysHisAspSerGlnAlaSerLysLeuGlyAlaProAlaAlaGlyGlyGlnGlnGln 140
DB 1067 CTTTGCATGAGCTCCAGAGCCAGTAAAGTTGGGGGCTCCGCGCCAGAGGGGCGAGAGAG 1126
QY 141 TrpGlnGlnGlnGlnArgGlnLeuGlnLysLysLysHisArgArgArgProSerLysLys 160
DB 1127 TGGGAGACAGACAGACAGCTGGGGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1186
QY 161 LysArgHisLysLysProTyrTyrLysLeuThrTrpGlnGlnGlnLysLysPheAspGln 180
DB 1187 AAGCGGCATTGGAGAACCGTACTCAAGCTGACCTGGAGAGAGAGAGAGAGAGAGAGAGAG 1246
QY 181 LysGlnSerLeuArgAlaSerArgLysArgAlaGlnMetPheAlaLysGlnGlnProVal 200
DB 1247 AAACAGAGCTTCAAGCTTCAAGATCCAGCGAGATGTTCCCAAGGGCCAGCGGCTC 1306

```

```

QY 201 AlaProTyrAsnThrThrGlnPheLeuMetAspAspHisAspGlnGlnGlnProAspLeu 220
DB 1307 GCGCCCTATTAACACACGAGAGTCTCTCATGATGATCAACAGACAGAGAGAGCGGATCTC 1366
QY 221 LysThrGlyLeuTyrSerLysArgAlaAlaAlaLysSerAspAspThrSerAspAsp 240
DB 1367 AAAACCGGCTGTACTCCAAAGCGGGCGCGCCCAATCCAGACACCGAGCATGACGAC 1426
QY 241 PheMetGlnGlnGlnGlyGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGln 260
DB 1427 TTCAATGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1486
QY 261 GluPheLeuGlnArgAspPheSerGlnThrTyrGlnArgTyrHisThrGlnSerLeuGln 280
DB 1487 GAGTTTCTGACAGCGGAGACTTCTCGAGAGCTACGAGCGGTACCAACGAGAGCGCTGAC 1546
QY 281 AsnMetSerLysGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGln 300
DB 1547 AACATGAGCAGACAGAGAGCTCATCAAGAGTACTCGAACTGAGAAAGTGCCTCTCGCGC 1606
QY 301 MetGluAspGluAsnAsnArgLeuArgLeuGlnGlnGlnGlnGlnGlnGlnGlnGlnGln 320
DB 1607 ATGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1666
QY 321 ArgValArgGlnLeuGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGln 340
DB 1667 CGTGTGCGGAGACTGTGAGCTGTGAGCTGTGAGCGGCTGCGCGGAGAGAGAGCTGCTG 1726
QY 341 ThrGluAsnGlnLeuHisArgGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGln 359
DB 1727 ACCGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1783

RESULT 2
US-10-205-823-171
/ Sequence 171, Application US/10205823
/ Publication No. US20030108963A1
/ GENERAL INFORMATION:
/ APPLICANT: Schlegel, Robert
/ APPLICANT: Monahan, John E.
/ APPLICANT: Endege, Wilson O.
/ APPLICANT: Gannavarapu, Manjula
/ APPLICANT: Gorbatcheva, Bella
/ APPLICANT: Hoersch, Sebastian
/ APPLICANT: Kamatkar, Shubhangi
/ APPLICANT: Monsey, Angela M.
/ APPLICANT: Glatt, Karen
/ APPLICANT: Zhao, Xumei
/ APPLICANT: Anderson, Dustin
/ TITLE OF INVENTION: NOVEL GENES, COMPOSITIONS, KITS, AND
/ TITLE OF INVENTION: METHODS FOR IDENTIFICATION, ASSESSMENT, PREVENTION, AND
/ TITLE OF INVENTION: THERAPY OF PROSTATE CANCER
/ FILE REFERENCE: MRI-044
/ CURRENT APPLICATION NUMBER: US/10/205,823
/ CURRENT FILING DATE: 2002-07-25
/ PRIOR APPLICATION NUMBER: 60/307,982
/ PRIOR FILING DATE: 2001-07-25
/ PRIOR APPLICATION NUMBER: 60/314,356
/ PRIOR FILING DATE: 2001-08-22
/ PRIOR APPLICATION NUMBER: 60/325,020
/ PRIOR FILING DATE: 2001-09-25
/ PRIOR APPLICATION NUMBER: 60/341,746
/ PRIOR FILING DATE: 2001-12-12
/ PRIOR APPLICATION NUMBER: 60/362,158
/ PRIOR FILING DATE: 2002-03-05
/ NUMBER OF SEQ ID NOS: 455
/ SOFTWARE: FastSeq for Windows Version 4.0
/ SEQ ID NO 171
/ LENGTH: 3624
/ TYPE: DNA
/ ORGANISM: Homo sapiens
US-10-205-823-171

```

Alignment Scores:  
 Pred. No.: 2,62e-164 Length: 3624  
 Score: 1910.00 Matches: 359  
 Percent Similarity: 100.00% Conservative: 0  
 Best Local Similarity: 100.00% Mismatches: 0  
 Query Match: 100.00% Indels: 0  
 DB: 15 Gaps: 0

US-09-972-758A-2 (1-359) x US-10-205-823-171 (1-3624)

```

QY 1 MetaIaGluProPheLeuSerGluTyrGlnHisGlnProGlnThrSerAsnCyThrGly 20
DB 690 ATGGCCGAGCCATCTTGTGAGATATCAACACACCTCAACTGACAGT 749
QY 21 AAlaAlaValGlnGluLeuAsnProGluArgProGluValaGluGluArgVal 40
DB 750 GCTGCTGCTGCTCCAGAAAGCTGAACCTGAGCCGCCCCAGGCGCGAGAGCGGGTG 809
QY 41 ProGluGluAspSerArgTyrGlnSerArgAlaPheProGlnLeuGlyArgProGly 60
DB 810 CCGAGAGAGAGACGTAGTGGCAATCGAGAGCGTCCCAAGTTGGTGCGCGGG 869
QY 61 ProGluGlyGluGlySerLeuGluSerGlnProProProLeuGlnThrGlnAlaCyPro 80
DB 870 CCGAGAGGGGAGAGAGCGCTGGAATCCCAACCTCCCTTGAGAGCCAGCGCTTCCA 929
QY 81 GluSerSerCysLeuArgGluGluGlyGlyGlyGlnAsnGlyAspAspSerSerAlaGly 100
DB 930 GAATCTAGCTGCTGAGAGAGGCGAGAAAGGCGCAAAATGGGAGACATGCTGCTGGC 989
QY 101 GlyAspPheProProProAlaGluValaGluProThrProGluValaGluLeuAlaGln 120
DB 990 GGGAGCTTCCCGCGCGGAGAAAGTGAACCGAGCCGAGGCGAGCGTCCGCCAG 1049
QY 121 ProCysHisAspSerGluAlaSerLysLeuGlyAlaProAlaAlaGlyGlyGluGlu 140
DB 1050 CCTTGTGATGACTCCAGAGCGCACTAAGTTGGGGGCTCCCGCCAGGGGCGAGAGAG 1109
QY 141 TrpGluGlnGlnArgGlnLeuGlyLysLysLysHisArgArgArgProSerLysLys 160
DB 1110 TGGGAGACAGAGAGAGAGCTGGGGAGAAACATAGAGAGAGCGCCGTCAGAAAG 1169
QY 161 LysArgHisTrpLysProTyrTyrLysLeuThrTrpGluGlyLysLysPheAspGlu 180
DB 1170 AAGCGCATTTGAAACCGTACTCAAGCTGACCTGGGAGAGAGAAAGTTCCAGCAG 1229
QY 181 LysGlnSerLeuArgAlaSerArgIleArgAlaGluMetPheAlaLysGlyGlnProVal 200
DB 1230 AAACAGAGCGCTTCAAGCTTCAAGATCCGAGCGGAGATGTTCCCAAGGGCCAGCGGTC 1289
QY 201 AlaProTyrAsnThrThrGlnPheLeuMetAspAspHisAspGlnGluGluProAlaPhe 220
DB 1290 GGGCCCTTAATACCAACGAGCTTCCATGAGATGATCAAGACAGAGAGAGCGGATCTC 1349
QY 221 LysThrGlyLeuTyrSerLysArgAlaAlaLysSerAspAspThrSerAspAspAsp 240
DB 1350 AAAACGGGCTGACTCCAGCGGGCGCGCCCAATCCAGACACACAGCATGACGAC 1409
QY 241 PheMetGluGluGlyGlyGluGluAspGlyGlySerAspGlyMetGlyGlyAspGlySer 260
DB 1410 TTCATGAGAAAGGGGTGAGAGATGGGGGCAAGATGGAGTGGAGGGGCGGACAGC 1469
QY 261 GluPheLeuGlnArgAspPheSerGluThrTyrGluArgTyrHisThrGluSerLeuGln 280
DB 1470 GAATTTCTGAGGGGAGCTTCTCGAGAGCGTACGAGCGGTACCAACGAGAGACCTGACG 1529
QY 281 AsnMetSerLysGlnGluLeuIleLysGlyTyrLeuGluLeuGluLysCysLeuSerArg 300
DB 1530 AACATAGCAGAGAGAGAGCTCATCAAGAGTACTGGAAGTGAAGAAAGTCCCTCGCC 1589
QY 301 MetGluAspGluAsnAsnArgLeuArgLeuGluSerLysArgLeuGlyGlyAspAspAla 320
DB 1590 ATGAGAGAGAGAGAAACCGGCTGGCTGGAGAGAGAGCGCTGGGTGGCGAGAGCGG 1649
  
```

```

QY 321 ArgValaArgGluLeuGluLeuGluLeuAspArgLeuArgAlaGlnAsnLeuGlnLeu 340
DB 1650 CGTGTGGCGAGAGCTGAGCTGAGCGGCTGAGCGGCGGAGAACTCCAGCTGCTG 1709
QY 341 ThrGluAsnGluLeuHisArgGlnGlnGluValaAlaProLeuSerLysPheGlyAsp 359
DB 1710 ACCGAGAAAGAACTGACCGGAGAGAGCGGCGGCTTCCAAAGTTTGGAGAC 1766
  
```

## RESULT 3

```

US-09-972-758-1
; Sequence 1, Application US/09972758
; Patent No. US20020160497A1
; GENERAL INFORMATION:
; APPLICANT: Case Western Reserve University
; APPLICANT: Montano, Monica
; APPLICANT: Wiltman, Bryan
; TITLE OF INVENTION: Suppressors of Human Breast Cancer Cell Growth
; FILE REFERENCE: 27708/04004
; CURRENT APPLICATION NUMBER: US/09/972,758
; PRIOR FILING DATE: 2001-10-05
; PRIOR APPLICATION NUMBER: US 60/238,187
; NUMBER OF SEQ ID NOS: 2
; SOFTWARE: PatentIn version 3.0
; SEQ ID NO 1
; LENGTH: 1080
; TYPE: DNA
; ORGANISM: Homo sapiens
US-09-972-758-1
  
```

Alignment Scores:  
 Pred. No.: 3,59e-164 Length: 1080  
 Score: 1902.00 Matches: 357  
 Percent Similarity: 99.72% Conservative: 1  
 Best Local Similarity: 99.44% Mismatches: 1  
 Query Match: 99.58% Indels: 0  
 DB: 10 Gaps: 0

US-09-972-758A-2 (1-359) x US-09-972-758-1 (1-1080)

```

QY 1 MetaIaGluProPheLeuSerGluTyrGlnHisGlnProGlnThrSerAsnCyThrGly 20
DB 1 ATGGCCGAGCCATCTTGTGAGATATCAACACACCTCAACTGACAGT 60
QY 21 AAlaAlaValGlnGluLeuAsnProGluArgProGluValaGluGluArgVal 40
DB 61 GCTGCTGCTGCTCCAGAAAGCTGAACCTGAGCCGCCCGAGGCGGAGAGCGGGTG 120
QY 41 ProGluGluAspSerArgTyrGlnSerArgAlaPheProGlnLeuGlyArgProGly 60
DB 121 CCGAGAGAGAGAGAGTGGGCAATCGAGAGCGTCCCACTGGTGGCGCTCGGG 180
QY 61 ProGluGlyGluGlySerLeuGluSerGlnProProProLeuGlnThrGlnAlaCyPro 80
DB 181 CCGAGAGGGGAGAGAGCGCTGAATCCCAACCTCCCTTGAGAGCCAGCGCTGTCCA 240
QY 81 GluSerSerCysLeuArgGluGlyGlyGlyGlnAsnGlyAspAspSerSerAlaGly 100
DB 241 GAATCTAGCTGCTGAGAGAGGCGAGAAAGGCGCAAAATGGGAGAGATCTGCTCGTGC 300
QY 101 GluAspPheProProProAlaGluValaGluProThrProGluValaGluLeuAlaGln 120
DB 301 GCGGACTTCCCGCGCGGAGAAAGTGAACCGAGCCCGAGAGCGGAGCTGCTCGCCAG 360
QY 121 ProCysHisAspSerGluAlaSerLysLeuGlyAlaProAlaAlaGlyGlyGluGlu 140
DB 361 CTTGTGATGATCCGAGGCGAGTAAAGTTGGGGCTCCGCGGAGGGGGCGAAGAGAG 420
QY 141 TrpGlyGlnGlnArgGlnLeuGlyLysLysLysHisArgArgArgProSerLysLys 160
DB 421 TGGGAGACAGCAGAGAGAGCTGGGAGAAACATTAAGAGAGCGCCGTCACAAAG 480
  
```



QY 161 LysArgHisTrpLysProTyrTyrLysLeuThrTrpGluLysLysPheAspGlu 180  
 Db 481 AACCGCATGAAACCGTACTCAAGCTGAACGCGGAGAAAGAAAGTTGACGAG 540  
 QY 181 LysGlnSerLeuArgAlaSerArgLleArgAlaGluMetPheAlaLysGlyGlnProVal 200  
 Db 541 AAACAGAGCCTTGAGCTTAAGATCCGAGCCAGATGTTCCCAAGGGCCAGCCGCTC 600  
 QY 201 AlaProTyrAsnThrTrpGlnPheLeuMetAspAspHisAspGlnGluGluProAspLeu 220  
 Db 601 GCCCCCTATAACACACGACGATTCCTCATGATGATACACACGAGAGAGCCGATCTC 660  
 QY 221 LysThrGlyLeuTyrSerLysArgAlaAlaAlaLysSerAspAspThrSerAspAspAsp 240  
 Db 661 AAAACCGCCTGTAATCTCAAGCCGCGCCGCAAAATCCAGACACACGATGACGAC 720  
 QY 241 PheMetGluGluGlyGlyGlyGluAspGlyGlySerAspGlyMetGlyGlyAspGlySer 260  
 Db 721 TTGATGAAAGAGGGGTGAGAGATGGGGGCGCATGGATGGGAGGGGACGGCAGC 780  
 QY 261 GluPheLeuGlnArgAspPheSerGluThrTyrGluArgTyrHisThrGluSerLeuGln 280  
 Db 781 GAGTTTTCGACGCGGACTCTTCGAGACGTAACGAGCGGTACACACGAGAGACCTTCAG 840  
 QY 281 AsnMetSerLysGlnGluLeuLleLysGlyTyrLeuGluLeuGluLysCysLeuSerArg 300  
 Db 841 AACATGAGCAAGCAGGAGCTCATCAAGAGTACTGGAACTGGAGAAAGTCTCTCGCGC 900  
 QY 301 MetGluAspGluMetAsnArgLeuArgLeuGluSerLysArgLeuGlyGlyAspAspAla 320  
 Db 901 ATGAGAGACAGAAACAACTGGCTGCGGCTGAGAGCAAGCGCTGGGTGGCGACGACGCG 960  
 QY 321 ArgValArgGluLeuGluLeuGluLeuAspArgLeuArgAlaGluMetLeuGlnLeuLeu 340  
 Db 961 CGCTGGGGGAGCTGAGAGCTGAGAGCTGACCGGCTGGCCCGAAGAACTCCACGCTGCTG 1020  
 QY 341 ThrGluAsnGluLeuHisArgGlnGlnGluArgAlaProLeuSerLysPheGlyAsp 359  
 Db 1021 ACCGAGAACGAACGTCACCGGACGAGAGAGCGCGCTTCCAAAGTTTGAGAGAC 1077  
 RESULT 4  
 US-10-264-237-629/C  
 ; Sequence 629, Application US/10264237  
 ; Publication No. US20040009491A1  
 ; GENERAL INFORMATION:  
 ; APPLICANT: Birse et al.  
 ; TITLE OF INVENTION: Nucleic Acids, Proteins, and Antibodies  
 ; FILE REFERENCE: P131PI  
 ; CURRENT APPLICATION NUMBER: US/10/264,237  
 ; CURRENT FILING DATE: 2002-10-04  
 ; PRIOR APPLICATION NUMBER: PCT/US01/16450  
 ; PRIOR FILING DATE: 2001-05-18  
 ; PRIOR APPLICATION NUMBER: US 60/205,515  
 ; PRIOR FILING DATE: 2000-05-19  
 ; NUMBER OF SEQ ID NOS: 2876  
 ; SOFTWARE: PatentIn Ver. 3.1  
 ; SEQ ID NO 629  
 ; LENGTH: 1083  
 ; TYPE: DNA  
 ; ORGANISM: Homo sapiens  
 ; FEATURE:  
 ; NAME/KEY: misc\_feature  
 ; LOCATION: (16)..(16)  
 ; OTHER INFORMATION: n equals a,t,g, or c  
 ; FEATURE:  
 ; NAME/KEY: misc\_feature  
 ; LOCATION: (27)..(27)  
 ; OTHER INFORMATION: n equals a,t,g, or c  
 ; FEATURE:  
 ; NAME/KEY: misc\_feature  
 ; LOCATION: (32)..(32)  
 ; OTHER INFORMATION: n equals a,t,g, or c  
 ; FEATURE:

; NAME/KEY: misc\_feature  
 ; LOCATION: (111)..(111)  
 ; OTHER INFORMATION: n equals a,t,g, or c  
 US-10-264-237-629  
 Alignment Scores:  
 Pred. No.: 1,06e-57 Length: 1083  
 Score: 733.00 Matches: 146  
 Percent Similarity: 98.65% Conservative: 2  
 Best Local Similarity: 98.65% Mismatches: 0  
 Query Match: 38,38% Indels: 1  
 DB: 12 Gaps: 0  
 US-09-972-758a-2 (1-359) x US-10-264-237-629 (1-1083)  
 QY 212 AspHisAspGlnGluGluProAspLeuLysThrGlyLysTyrSerLysArgAlaAlaAla 231  
 Db 1081 GATCAGACCAAGAGAGCGGAGCTCTCAAAACCGGCTTACTTCCAAACGSSCGCCGC- 1023  
 QY 232 LysSerAspAspThrSerAspAspAspPheMetGluGluGlyGlyGlyGluAspGlyGly 251  
 Db 1022 AAATCCGACGACACACGCGATGACGACTTCATGAAAGAGGGGTGAGAGAGTGGGGC 963  
 QY 252 SerAspGlyMetGlyGlyLysArgLysGluPheLeuGlnArgAspPheSerGluThrTyr 271  
 Db 962 AGCGATGGGATGGAGGGGACGCGACGCGAGTTTCTGCAAGCGGACTTCTCGAGACGTAC 903  
 QY 272 GluArgTyrHisThrGluSerLeuGlnAsnMetSerLysGlnGluLeuLleLysGlyTyr 291  
 Db 902 GAGCGGTACCAACGAGAGAGCTTCGACAACTAGACACAGAGAGTCTTCAAGAGACTAC 843  
 QY 292 LeuGluLeuGluLysCysLeuSerArgMetGluAspGluAsnAsnArgLeuArgLeuGln 311  
 Db 842 CTGGAACCTGAGAAAGTCCCTTCGCGCATGAGAGACGAAACAACCGGCTGCGCTGAG 783  
 QY 312 SerLysArgLeuGlyGlyLysAspAlaArgValArgGluLeuGluLeuLeuAspArg 331  
 Db 782 AGCAAGCGGCTGGTGGCCACGACGCGCTGTGCGGAGACTGGAGCTGGAGCTGGACCGG 723  
 QY 332 LeuArgAlaGluAsnLeuGlnLeuLeuThrGluAsnGluLeuHisArgGlnGlnArg 351  
 Db 722 CTGCGCGCCGAGAACTCCAGCTGCTGACCGCAAGAACTGCAACCGGACGAGAGCGA 663  
 QY 352 AlaProLeuSerLysPheGlyAsp 359  
 Db 662 GCGCGCTTCCAAAGTTTGAGAGAC 639  
 RESULT 5  
 US-09-983-965-4895  
 ; Sequence 4895, Application US/09983965  
 ; Patent No. US20020137160A1  
 ; GENERAL INFORMATION:  
 ; APPLICANT: Warren, Wesley C.  
 ; APPLICANT: Tao, Nengbing  
 ; APPLICANT: Byatt, John C.  
 ; APPLICANT: Mathalagan, Nagappan  
 ; TITLE OF INVENTION: NUCLEIC ACID AND OTHER MOLECULES ASSOCIATED WITH LACTATION AND  
 ; FILE REFERENCE: 37-21(1029)C  
 ; CURRENT APPLICATION NUMBER: US/09/983,965  
 ; CURRENT FILING DATE: 2001-10-26  
 ; PRIOR APPLICATION NUMBER: US 09/465,231  
 ; PRIOR FILING DATE: 1999-12-15  
 ; PRIOR APPLICATION NUMBER: US 60/113,678  
 ; PRIOR FILING DATE: 1998-12-17  
 ; NUMBER OF SEQ ID NOS: 5912  
 ; SEQ ID NO 4895  
 ; LENGTH: 414  
 ; TYPE: DNA  
 ; ORGANISM: Bos taurus  
 ; FEATURE:  
 ; OTHER INFORMATION: Clone ID: 24-LIB34-008-Q1-E1-F7  
 US-09-983-965-4895

Alignment Scores:	
Pred. No.:	2,39e-54
Score:	691.00
Percent Similarity:	95.62%
Best Local Similarity:	94.89%
Query Match:	36.18%
DB:	10
Length:	414
Matches:	130
Conservative:	1
Mismatches:	6
Indels:	0
Gaps:	0

US-09-972-758A-2 (1-359) x US-09-983-965-4895 (1-414)

QY	155	ATGATGProterelyslyslslyslsATGhstPLsPSPProTYrLYzLeuThrTrgIUGlu	174
Db	2	AGAGCCCCCTCCAAAGAAAGAACGGCTTGGAAACCTACTATACCTCAGCTGGAGAG	61
QY	175	LYsLYsLYsPheAspGLuLYsgIInserLeuATGAlAserArgIeArgAlGLuMetPhe	194
Db	62	AAGAAAAAGTTTCATGATGAAAAACAGACCTCGCGAGTTTCAGAGATTCCAGCGAGATGTTCC	122
QY	195	AlAlAsGLyGLInProValAlAPoLYrAenThrThrGLInPheLeuMetAspAspHisAsp	211
Db	122	GCCAAAGGCCAGCCAGCTGCTCCCTTAAACAACGACAGCTTCATAGATGACACACAC	183
QY	215	GLInGLInLProAspLeuLYserThrGLyLeuLYrSerLYsArgAlAlAlLYsSerAsp	234
Db	182	CAGGAGGAGCCGGATCTTTAAACCGGCTCTTATCCAAAGGGCGCGCTGCCAATCCGAC	24
QY	235	AspThrSerAspAspAspPheMetGLuGLuGLyGLyGLuGLuAspGLyGLySerAspGLy	255
Db	242	GACCCAGCGCATGAGGACTTTATGGAAGAAAGCGGGGAGAGAGATGGGGGCGACGACGGG	303
QY	255	MetGLyLYsAspGLySerGLuPheLeuGLInATGAspPheSerGLuThrTYrGLuArgTYr	276
Db	302	ATGGGAGGAACGGCGACGAGTTTTCGACGGCGGACTTCTCGGAACAACTATGAGCGGTAC	361
QY	275	HisThrGLuSerLeuGLInAsenMetSerLYsgInGLuLeuILeLYsGLuTYr	291
Db	362	CACCGGAGAGCCTGCAGAACATGACCAAGCAGAGACTTATCAAGAGATAC	412

## RESULT 6

```

US-09-954-531-813
: Sequence 813, Application US/09954531
: Patent No. US20020165180A1
: GENERAL INFORMATION:
: APPLICANT: Weaver, Zoe
: TITLE OF INVENTION: Process for Identifying Anti-Cancer Therapeutic Agents Using Cancer
: TITLE OF INVENTION: Gene Sets
: FILE REFERENCE: 689290-77
: CURRENT APPLICATION NUMBER: US/09/954,531
: CURRENT FILING DATE: 2002-05-02
: PRIOR APPLICATION NUMBER: US/60/233,133
: PRIOR FILING DATE: 2000-09-18
: PRIOR APPLICATION NUMBER: US/60/234,009
: PRIOR FILING DATE: 2000-09-20
: PRIOR APPLICATION NUMBER: US/60/234,034
: PRIOR FILING DATE: 2000-09-20
: PRIOR APPLICATION NUMBER: US/60/234,509
: PRIOR FILING DATE: 2000-09-22
: PRIOR APPLICATION NUMBER: US/60/234,567
: PRIOR FILING DATE: 2000-09-22
: NUMBER OF SEQ ID NOS: 1392
: SOFTWARE: PatentIn version 3.0
: SEQ ID NO 813
: LENGTH: 461
: TYPE: DNA
: ORGANISM: Homo sapiens
: FEATURE:
: NAME/KEY: misc_feature
: LOCATION: (1)..(461)
: OTHER INFORMATION: n=a,t,g or c
US-09-954-531-813

```

Pred. No.:	1,2e-48	Length:	461
Score:	629.00	Matches:	120
Percent Similarity:	93.85%	Conservative:	2
Best Local Similarity:	92.31%	Mismatches:	8
Query Match:	32.93%	Indels:	1
DB:	10	Gaps:	0

US-09-972-758A-2 (1-359) X US-09-954-531-813 (1-461)

QY	1	MetAlaGluProPheLeuSerCyluYrGlnIsgInProGlnThrSerAsnCySthGly	20
Db	72	ATGCGCGAAGCCATTCTTGTCAGAAATATACAACCCAGCTCAAACTAGCACTGACAGGT	131
QY	21	AlAlaAlaValGlnGluGluLeuAsnProGlnuArgProProGluYalAgluGluArgVal	40
Db	132	GGTGGTGGTGTCCAGGAAGAGCTGAACCTTAGGGCCCCCAGGCGNACGAGCGGGTG	191
QY	41	ProGluGluAspSerArgTrpGlnSerArgAlaPheProGlnLeuGlyGlyArgProGly	60
Db	192	CCGAGAGAGACAGATGAGTGGCAATCGAGAGCGTTCCCCAGTTGGGGGCGGTCGCGGG	251
QY	61	ProGluGlyGluGlySerLeuGluSerGlnProProProLeuGlnThrGlnAlaCysPro	80
Db	252	CCGAGACGGGGAAGGAGCGCTGGAATCCCAACCACTCTCTTGACAGCCAGCCTGTCCA	311
QY	81	GluSerSerCysLeuArgGluGlyGluYulYsgGlyGlnbncGlyAspAspSerSerAlaGly	100
Db	312	GAATCTAGCTGCTCTAGAGAGGGCCAGAAAGGGCCAGAAATGGGAGCACTGTCGGCTGGC	371
QY	101	GlyAspPheProProProAlaGluValGlnProThrProGluAlaGluLeuLeuAlaGln	120
Db	372	GC-GAATTCGCCGCGNCGCAGAAAGTGAAGCCGACGCCGAGGCGGAGTGCCTTGCACAG	430
QY	121	ProCysHisAspSerGluAlaSerYleu	130
Db	431	CCTTGTCATGACTCCGAGGGCAGTAAAGTTG	460

## RESULT

```

US-09-918-995-403/c
Sequence 403, Application US/09918995
Publication No. US2003007362A1
GENERAL INFORMATION:
APPLICANT: Hyseq, Inc.
TITLE OF INVENTION: NOVEL NUCLEIC ACID SEQUENCES OBTAINED
FROM VARIOUS CDNA LIBRARIES
FILE REFERENCE: 20411-756
CURRENT APPLICATION NUMBER: US/09/918,995
CURRENT FILING DATE: 2001-07-30
PRIOR APPLICATION NUMBER: US-09/235,076
PRIOR FILING DATE: 1999-01-20
NUMBER OF SEQ. ID NOS: 38054
SOFTWARE: FastSeq for Windows Version 3.0
SEQ ID NO 403
LENGTH: 495
TYPE: DNA
ORGANISM: Homo sapiens
FEATURE:
NAME/KEY: misc_feature
LOCATION: (1)...(495)
OTHER INFORMATION: n = A,T,C or G
US-09-918-995-403

```

Pred. No.:	1.06e-47	length:	499
Score:	619.00	Matches:	120
Percent Similarity:	98.37%	Conservative:	0
Best local Similarity:	98.37%	Mismatches:	2
Query Match:	32.41%	Indels:	0
DB:	11	Gaps:	0

US-09-972-758A-2 (1-359) X US-09-918-995-403 (1-495

QY 237 SerAspAspPheMetGluGluGlyGlyGluGluAspGlyGlySerAspGlyMetGly 256

```

Db      495 AGCATATACACATTCATGAGAGAGGCGGTGAGAGCATGGGCGACCGATGGATGGGA 436
Qy      257 G1yAaspGlySerGluPheLeuGluAraArgPheSerGluThrTyrgLysArgHisThr 276
Db      435 GGGCAGCGGACCGGATTTCTTCACCGGGAATCTTCGAGAGACGTAACGCGCTTACACAG 376
Qy      277 GluSerLeuGluAsnMetSerLysGlnGluLeuLysGluTyrgLysGluLeuGluLys 296
Db      375 GAGAGCTTGCAAGACATGAGACACGAGCTCATCAAGAGTACTCTGGAACCTGGAGAG 316
Qy      297 CysLeuSerArgMetGluAerGluAerGluAerGluAerGluAerGluAerGluAerGlu 316
Db      315 TGCTCTCGCGCATGAGAGAGAGAGACACCGGCTGCGCTGAGAGACACCGCTGGGT 256
Qy      317 G1yAaspAraValArgValArgGluLeuGluLeuGluLeuAraArgLysArgLysArg 336
Db      255 GGGCAGCAGCGCGCTGCGGAGAGCTGAGCTGAGCTGAGCTGAGCTGAGCTGAGAGAG 196
Qy      337 LeuGluLeuLeuThrGluAerGluLeuHisArgGlnGlnGlnLysArgLysArgLys 356
Db      195 CTCAGCTGCTGACCGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 136
Qy      357 PheGlyAer 359
Db      135 TTGGAGAGC 127

```

## RESULT 8

```

US-10-094-749-1223
/ Sequence 1223, Application US/10094749
/ Publication No. US20030219741A1
/ GENERAL INFORMATION:
/ APPLICANT: ISOGAI, TAKAO
/ APPLICANT: SUGIYAMA, TOMOYASU
/ APPLICANT: OTSUKI, TETSUJI
/ APPLICANT: WAKAMATSU, AI
/ APPLICANT: SATO, HIROYUKI
/ APPLICANT: ISHII, SHIZUKO
/ APPLICANT: YAMAMOTO, JUN-ICHI
/ APPLICANT: ISONO, YUUKO
/ APPLICANT: HIO, YURI
/ APPLICANT: OTSUKA, KAORU
/ APPLICANT: NAGAI, KEIICHI
/ APPLICANT: IRIE, RYOTARO
/ APPLICANT: TAMECHIKA, ICHIRO
/ APPLICANT: SEKI, NAOHICO
/ APPLICANT: YOSHITAKA, TSUTOMU
/ APPLICANT: OTSUKA, MOTOTUKI
/ APPLICANT: NAGAHARI, KENJI
/ APPLICANT: MASUHO, YASUHIKO
/ TITLE OF INVENTION: NOVEL FULL-LENGTH CDNA
/ FILE REFERENCE: 084335/0160
/ CURRENT FILING DATE: 2002-03-12
/ PRIOR APPLICATION NUMBER: 60/350,435
/ PRIOR FILING DATE: 2002-01-24
/ PRIOR APPLICATION NUMBER: JP 2001-328381
/ PRIOR FILING DATE: 2001-09-14
/ NUMBER OF SEQ ID NOS: 3381
/ SOFTWARE: Patent Ver. 2.1
/ SEQ ID NO 1223
/ LENGTH: 1330
/ TYPE: DNA
/ ORGANISM: Homo sapiens
US-10-094-749-1223

```

## Alignment Scores:

```

Pred. No.: 1.58e-41 Length: 1330
Score: 556.50 Matches: 141
Percent Similarity: 51.40% Conservative: 42
Best Local Similarity: 39.61% Mismatches: 118
Query Match: 29.14% Indels: 55
DB: 13 Gaps: 9

```

```

US-09-972-758a-2 (1-359) x US-10-094-749-1223 (1-1330)
Qy      20 G1yAlaAlaValGlnGluLeuAerProGluAraArgProProGlyAlaGluGluAra 39
Db      87 GAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 146
Qy      40 ValProGluGluAerSerArgTyrGlnSerArgAlaPheProGlnLeuGluGlyArgPro 59
Db      147 -----GAAACAGAGGTGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 185
Qy      60 GlyProGluGluGluGluSerLeuGluSerGln----- 70
Db      186 -----GAGAAATAGAGGTCTCACTATTCAGAGAGAGAGAGAGAGAGAGAGAGAG 239
Qy      71 -----ProProProLeuGlnArgLysArgLysArgLysArgLysArgLysArgLys 88
Db      240 ATGATGAGCAGCTCCAGACAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 299
Qy      89 G1yLysGlyGlnAerGlyAerAerSer-----SerAlaGly 100
Db      300 GCGAAGAGCTCTGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 359
Qy      101 G1yAerPheProProProAlaGluValGluProThrProGluAlaGluLeuAerGln 120
Db      360 GGTTCCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 419
Qy      121 ProCys-----HisAerSerGluAerLysArgLysArgLysArgLysArgLysArg 136
Db      420 GCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 479
Qy      137 G1yGluGluGluTyrPylGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGln 156
Db      480 TCAGCGAG-----GCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 521
Qy      157 ProSerLysLysLysArgHisTyrLysProTyrTyrLysLysLysLysLysLysLys 176
Db      522 CCATGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 581
Qy      177 LysPheAerGluLysGlnSerLeuAerArgLysArgLysArgLysArgLysArgLys 196
Db      582 CAGCGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 641
Qy      197 G1yGlnProValAlaProTyrAerThrThrGlnPheLeuMetAerAerAerAerAer 216
Db      642 GGGCAGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 701
Qy      217 GluPro-----AerLeuLysThrGlyLeuTyrSerLysArgAlaAlaLysSerAer 234
Db      702 GAGCCCAACTTGATGTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 758
Qy      235 AspThrSerAerAerAerPheMetGluGluGluGluGluGluGluGluGluGluGlu 254
Db      759 -----GAGCGCGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 779
Qy      255 MetGlyGlyAerGlySerGluPheLeuGlnAraArgPheSerGluThrTyrgLysArg 274
Db      780 CCGGCGCGAGCGCGAGAGTTCAGCGGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 839
Qy      275 HisThrGluSerLeuGlnAerMetSerLysGlnGluLeuLysGluTyrgLysGluLeu 294
Db      840 CACACCGAGAGAGCTGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 899
Qy      295 G1yLysCysLeuSerArgMetGluAerGluAerAerAerAerAerAerAerAerAer 314
Db      900 GAGAGAGCGCTGCTGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 959
Qy      315 LeuGlyGlyAerAerAerAerAerAerAerAerAerAerAerAerAerAerAerAer 333
Db      960 ACCGCGCAGAGCTCTGCGCGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1019
Qy      334 AlaGluAerLeuGluLeuThrGluAerGlnGlnGlnGlnGlnGlnGlnGlnGlnGln 349

```

DB 1020 ACCGAAACCGAGCGCTTGTCTGACGAGAACGATGTGACCGAGAG 1067

RESULT 9  
US-10-108-260A-1496/c  
Sequence 1496, Application US/10108260A  
Publication No. US20040005560A1  
GENERAL INFORMATION:  
APPLICANT: HELIX RESEARCH INSTITUTE  
TITLE OF INVENTION: No. US20040005560A1e1 full length cdna  
FILE REFERENCE: H1-A0106  
CURRENT APPLICATION NUMBER: US/10/108, 260A  
NUMBER OF SEQ ID NOS: 5458  
SOFTWARE: PatentIn Ver. 2.1  
SEQ ID NO 1496  
LENGTH: 2048  
TYPE: DNA  
ORGANISM: Homo sapiens  
US-10-108-260A-1496

Alignment Scores:  
Pred. No.: 2, 86e-40 Length: 2048  
Score: 545.00 Matches: 132  
Percent Similarity: 55.77% Conservative: 42  
Best Local Similarity: 42.31% Mismatches: 95  
Query Match: 28.53% Indels: 43  
Gaps: 7

DB: 12

US-09-972-758A-2 (1-359) x US-10-108-260A-1496 (1-2048)

QY 54 GlnLeuGlyGlyArgProGlyProGlyGlyGlyGlySerLeuGlySerLeu----- 70  
DB 1929 AAGGTAGGCTTCAAGCTAGGTAGAGGAGAAAGTTCCACTCCAGAGACAGCTGTCTG 1870  
QY 71 -----ProProLeuGlnThrGlnAlaCysProGlySerSerSerSerSer 85  
DB 1869 ATCCATCTTCTTCTCTCTCTCTCTCTCTTCTTCTCTCTCTCTCTCTCTCTCTCTCT 1824  
QY 85 uArgGlyGlyGlyGlyGlyGlyGlyGlyGlyGlyGlyGlyGlyGlyGlyGlyGlyGlyGly 105  
DB 1823 -----GGAGGCCCAACACCCCTGAGCTCATGACTGTGAGTGTCTCTCTCTCTCTCT 1771  
QY 105 oProAlaGlyGlyGlyGlyGlyGlyGlyGlyGlyGlyGlyGlyGlyGlyGlyGlyGlyGly 122  
DB 1770 GACACCGCGGATGAGAGCCACCTAGAGTGAAGTCTTCTGCGGGCTGTGCGTGGCT 1711  
QY 123 ---HisAspSerGlnAlaSerLeuGlyAlaProAlaAlaGlyGlyGlyGlyGlyGly 141  
DB 1710 GGGCTGGAACAGTACGAGTCCCGGACCCAGAGCCGAGGCGCTCTCAAGCGAG----- 1656  
QY 141 pGlyGlnGlnGlnAlaGlyGlyGlyGlyGlyGlyGlyGlyGlyGlyGlyGlyGlyGlyGly 161  
DB 1655 -----GCTGTGCTGGCCCGGAGAAACCGTGGCGGCGCATGAGCGCAA 1609  
QY 161 sArgHisTrpLeuProGlyGlyGlyGlyGlyGlyGlyGlyGlyGlyGlyGlyGlyGlyGly 181  
DB 1608 AAGGCACTGGCGACCTCACTGAGCTGAGCTGGCTGAGAAACACACAGCGGATAGAG 1549  
QY 181 sGlnSerLeuArgAlaSerArgGlyGlyGlyGlyGlyGlyGlyGlyGlyGlyGlyGlyGly 201  
DB 1548 GCAGACCCAGAGGCGCTCCCGGCTCCCGGAGAGATGTTCCCAAGAGCGCAAGCCCTGGC 1489  
QY 201 aProGlyArgTrpThrGlnPheLeuMetAspAspArgHisAspGlnGlnPro-----As 219  
DB 1488 CCCCACCAACACACCACTTCTGATGATGACAGGACCCGAGAGAGCCCAACTTGA 1429  
QY 219 pLeuGlyThrGlyLeuGlySerArgAlaAlaAlaAlaSerSerAspAspThrSerSerAsp 239  
DB 1428 TGtGCCCATGGATCTCCACACCGAGTTCCAGTGGGAGAGT----- 1386  
QY 239 pAspPheMetGlyGlyGlyGlyGlyGlyGlyGlyGlyGlyGlyGlyGlyGlyGlyGlyGlyGly 259  
DB 1385 -----GAGCGCGGGGACAGTATGGCGGGGCGAGCGCA 1351

QY 259 ySerGlnPheLeuGlnArgAspPheSerGlnThrThrGlyGlyArgGlyArgHisThrGlySer 279  
DB 1350 CGGTAGTTCACGCGAAGAGACTTCTGTAGACTTACGAACTCTCCACACGAGAGCTT 1291  
QY 279 uGlnAspMetSerLeuGlnGlnLeuLeuLeuGlyGlyGlyGlyGlyGlyGlyGlyGlyGly 299  
DB 1290 GAGAGCCCGACAGCAAGAGAGAGTGTGTGAGACTTCTGTGAGAGAGAGAGAGAGAGCTTCT 1231  
QY 299 rArgMetGlnAspGlnAspAsnArgLeuArgLeuGlySerLeuGlyGlyGlyGlyGlyGly 319  
DB 1230 GCAGCGGAG 1171  
QY 319 pAlaArg---ValArgGlyLeuGlyGlyGlyGlyGlyGlyGlyGlyGlyGlyGlyGlyGlyGly 338  
DB 1170 CTGCCCGCAGGTGAG 1111  
QY 338 nLeuLeuThrGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGln 349  
DB 1110 GCTTGTCTGAGAGAACAGATGTGAGACCGAGAG 1077

RESULT 10  
US-09-991-936-1712  
Sequence 1712, Application US/09991936  
Publication No. US20030073827A1  
GENERAL INFORMATION:  
APPLICANT: Brandt, Kevin S.  
APPLICANT: Gaines, Patrick J.  
APPLICANT: Stinchcomb, Dan T.  
APPLICANT: Wisniewski, Nancy  
TITLE OF INVENTION: FLEA HEAD, NERVE CORD, HINDGUT AND MALPIGHIAN TUBULE  
FILE REFERENCE: FC-6-C1  
CURRENT APPLICATION NUMBER: US/09/991, 936  
PRIOR FILING DATE: 2001-11-21  
PRIOR APPLICATION NUMBER: US/09/543, 668  
PRIOR FILING DATE: 2000-04-07  
PRIOR APPLICATION NUMBER: 60/128, 704  
NUMBER OF SEQ ID NOS: 1959  
SOFTWARE: PatentIn Ver. 2.1  
SEQ ID NO 1712  
LENGTH: 549  
TYPE: DNA  
ORGANISM: Ctenocephalides felis  
US-09-991-936-1712

Alignment Scores:  
Pred. No.: 4, 91e-08 Length: 549  
Score: 184.00 Matches: 66  
Percent Similarity: 42.29% Conservative: 30  
Best Local Similarity: 29.07% Mismatches: 53  
Query Match: 9.63% Indels: 78  
Gaps: 12

DB: 11

US-09-972-758A-2 (1-359) x US-09-991-936-1712 (1-549)

QY 34 ProGlyAlaGlyGlyGlyGlyGlyGlyGlyGlyGlyGlyGlyGlyGlyGlyGlyGlyGly 52  
DB 33 CCAGCGGTGAGAAATATG-----AGTAATATATCAATCAAAAGTGTTTTAA 80  
QY 53 ---ProGlnLeuGlyGlyArgProGlyProGlyGlyGlyGlyGlyGlyGlyGlyGlyGly 71  
DB 81 TTAATACTGTTAACATCAAAATGAGTGTAAAAATTGAA---AATTTGAATCGGAAAT 137  
QY 72 Pro-----ProLeuGlnThrGlnAla 78  
DB 138 CCAAAATATATCAACCCGAGTACAGCGTCAAAATGTCTGCCCTTCAACAGCGCT 197  
QY 79 CysProGlySerSerCysLeuArgGlyGlyGlyGlyGlyGlyGlyGlyGlyGlyGlyGlyGly 98  
DB 198 -----ACTGAAGCGCCTCTATCAGAGAAAAAGAAATGATATGAT----- 239

```

QY 99 AlAGlyGlyAspPheProProAlaGluValGluProThrProGluAlaGluLeu 118
Db 239 -----
QY 119 AlAGlnProCysHisAspSerGluAlaSerLysLeuGlyValProAlaAlaGlyGlu 138
Db 240 -----AGAAAAAGAAAGCAACCAAGTGGT-----GCACCAAAATGGGAG 278
QY 139 GlUGlUtrpGlyGlnGlnGlnAArgGlnLeuGlyLysLysLysHisArgArgProSer 158
Db 279 -----AAGAAAAAAGCTAGACAGCAAGTAATGT 305
QY 159 LysLysLysArgHisArgLysProTyrrTyrrLysLeuThrTrpGluGluLysLysPhe 178
Db 306 AAGAGAAAACTAAT---AAACCGTACATATAAACACCTGGACCCCACTGAAGAATGTC 362
QY 179 AspGluLysGlnSerLeuAlaGlyAlaSerArgLysGlyAlaGluMetPheAlaLysGlyGln 198
Db 363 ---CAAAAGAAATGAGAGAGTTCGTAGTCCCGCTAAATATTATTAAGCATGGGGCAC 419
QY 199 ProValAlaProTyrrAenThrThrGlnPheLeuMetAspAspHisAspGlnGluPro 218
Db 420 ACATTAGTCTCTGTAACAGAACCAATTTCTCATGGAAGATCATGAT----- 467
QY 219 AspLeuLysThrGlyLeuTyrrSerLysArgAlaAlaLysSerAspAspThr----- 236
Db 468 -----GTCCTCACCAAGATTCTCATCTGCAGACCTGCAGCTCTACTTAGTGT 512
QY 237 ---SerAspAspAspPheMet 242
Db 513 CGTTCTGAAGACACTCTMATG 533

RESULT 11
US-10-084-817-158
/ Sequence 158, Application US/10084817
/ Publication No. US20030119009A1
/ GENERAL INFORMATION:
/ APPLICANT: Susan Stuart
/ APPLICANT: Jed G. Nuchtern
/ APPLICANT: Sharon E. Pion
/ APPLICANT: Jason M. Shohet
/ TITLE OF INVENTION: GENES REGULATED BY MYCN ACTIVATION
/ FILE REFERENCE: PA-0046 US
/ CURRENT APPLICATION NUMBER: US/10/084,817
/ PRIOR FILING DATE: 2002-02-25
/ PRIOR APPLICATION NUMBER: 60/270,784
/ NUMBER OF SEQ ID NOS: 365
/ SOFTWARE: PERL Program
/ LENGTH: 6354
/ TYPE: DNA
/ ORGANISM: Homo sapiens
/ FEATURE:
/ NAME/KEY: misc feature
/ OTHER INFORMATION: Incyte ID No. US20030119009A1 1393778CB1
US-10-084-817-158

Alignment Scores:
Pred. No.: 7.8e-05 Length: 6354
Score: 162.00 Matches: 99
Percent Similarity: 35.73% Conservative: 65
Best Local Similarity: 21.57% Mismatches: 183
Query Match: 8.48% Indels: 113
DB: 15 Gaps: 15

US-09-972-758a-2 (1-359) x US-10-084-817-158 (1-6354)
QY 2 AlAGlnProCysHisAspSerGluLysGln-----HisGlnProGln--- 14
Db 2176 GGTGATGGCTACGCTGAGAACACCAACCCCAACTTGTCCGCTGCTATCCCAACCA 2235
QY 15 -----ThreAsnCysThrGlyAlaAlaValGlnGlu 26

```

```

Db 2236 CGAAGAAAGCCCGGACGTGAGACCCTGCATCTCGTGTGAGACCAAGCTCGCTGCAACCG 2295
QY 27 -----GluLeuAsnProGluArgPro-----ProGly 35
Db 2296 TGTTCCTGAGGGCATCCGTATCTGCCGCCAGGGCTTCCCAACAGGGTGTCTTCCAGGA 2355
QY 36 AlAGluGluArgValProGluGluAspSerArgTrpGlnSerArgAlaPhe-----Pro 53
Db 2356 GTTTGGCAGAGATATGATGATCTGATCTCCAAATCCATCTCCAAAGGGTTTATGAGACG 2415
QY 54 GlnLeuGlyGlyArgProGlyProGlyGlyGlySerLeuGlnSerGlnProProPro 73
Db 2416 GAAGCAGGGGTGGTCTCATGATTAAGCCCTGGAGCTCGACAGCAATCTGTACCGCAN 2475
QY 74 LeuGlnThrGlnAlaCysProGlnSerSerCys----- 84
Db 2476 TGGCCAGACCAAGTCTTTCTTCCTCCGTGCGGTGCTGCGCCACCTGGAGAGAGCGACA 2535
QY 85 -----LeuArgGluGlyLysGlyGln 92
Db 2536 CCGAAGATCACCGACGTATCATAGGTTCCAGGCTGCTGAGGGGCTTACCTGGCCAG 2595
QY 93 AsnGlyAspAspSerSerAlaGlyLysPhe-----ProProProAlaGluVal 109
Db 2596 GAAAGCATTTGCCAAGCCGACAGCAGCACTTACCCCATGAAGTCTTCAGGCGAACTG 2655
QY 110 GluProThrProGluAlaGluLeuLeuAla-----GlnProCysHisAspSerGluAla 127
Db 2656 CGCTGCTTACTGAGAGCTGCGAATCGGACGTGGTGGCCGCTTTCACCAAGTCAAGCC 2715
QY 128 SerLysLeuGlyAlaProAlaAlaGlyGlyGlnGluLysTrp-----GlyGln 143
Db 2716 GCTGCTGACAGGTGAGCGCGGACAGAGAGATGATGCGCAAGAGAGAGAGCTGTGA 2775
QY 144 GlnGlnArgGln-----LeuGlyLysLysLysHisArgArgProSerLysLys 160
Db 2776 GGTCAAGAGAAAGCAGCTGCTGCGAGAACAGGCTCACGAGATGAGACGCTGCAGTTC 2835
QY 161 LysArgHisTrp-LysProTyrrTyrrLysLeuThrTrpGluGluLysLysPheAsp 180
Db 2836 TCAGCTCATGCGACAG-----AAATTGACGTGCGAGAGAGCTTCAGGCGAAGAAC 2886
QY 180 LysGlnSerLeuArgAlaSerArgLysArgLysLeuMetPheAlaLysGlyGlnProVa 200
Db 2887 CGAGCTGTGCGCAGAGCTGAGAGAGCTCCGGCCCTGACCGCCAAAGAGAGAAATT 2946
QY 200 AlaPro----- 202
Db 2947 AGAAGATCTGCATGACCTAGAGGCGGAGGAGAGAGAGAGAGAGAGAGAGAGAGAG 3006
QY 203 -----TyrAsnThrThrGlnPheLeuMetAspAspHis 214
Db 3007 CCGCAGCGCGAGAGAGAGAGATGACAGCAAACTCCAGAGCTTGAGAGAGAGAGAGAG 3066
QY 214 pGlnGluGluProAspLeuLysThrGlyLeuTyrrSerLysArgAlaAlaLysSerAs 234
Db 3067 GAGAGAGAG-----AGCCCGCGCAGAGCT- GCGCTGCGAGAGAGCTA 3110
QY 234 pAspThrSerAspAspAspPheMetGlnGlyGlyGlyGluAspGlySerAspG 254
Db 3111 CCACCGAGCGAAGCTGAAGAAAGCTGAGAGAGAGAGAGATCTCTCGAGAGAGAGAGACT 3170
QY 254 yMetGlyGlyAspGlySerGlnPheLeuGlnArgAspPheSerGlnThrTyrrGluArg 274
Db 3171 GCAAGCTGCGCAAGAAAGAACTGCTGGAAGACAGAAATACGTGATTCACCAACACC 3230
QY 274 rHis-ThrGlnSerLeuGlnAsnMetSerLysGlnGluLeuLeuLysGlyTrp----- 291
Db 3231 TCACAGAAAGAGAGAGAAATCTAAGAGCTGCGCAAGCTGCAAGAAAGAGAGAGAGAG 3290
QY 292 -----LeuGluLeuGlyLysCysLeuSerArgMetGluAspGluAsnAsnArgLeuArg 310

```



```

QY 259 yserglupheleuglarapspeserluthrtyrgluarqlyrhie-thrcluserl 279
DB 3117 AAAAGAACTGCTGTAAGACAGATAGCTGAGTTCAACCACTCAGAGAGAGAG 3176
QY 279 euglnamhetserlysglngluileuileysgluyr-----leugluileug 295
DB 3177 AGAAATCTAAGAGCTTCGCCAGCAAGCAAGATGAGCAATGATCACTGACTTGG 3236
QY 295 lulyesCyseuserArmetgluapglubnbnanargluarqlyseugluserlyargl 315
DB 3237 AAGAGCGCCTCCCGCAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 3296
QY 315 eucllyllyap-----AspAlaryValArg 324
DB 3297 TGGAGGAGAGCTCCACAGACTTCAGCAGCAGCAGCTCCAGGCCAGCATCGCG 3356
QY 324 luleugluileugluileuapbarygluarqlyaglulbnleugluileuthgluabng 344
DB 3357 AGCTCAAGATGCGCTGGCCAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 3416
QY 344 luleuhsArqnglngluarqlyAProleuserlyr 356
DB 3417 AAGAGAGAGCTGCCAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 3454

```

```

RESULT 13
US-10-107-782-35
Sequence 35, Application US/10107782
Publication No. US20040018970A1
GENERAL INFORMATION:
APPLICANT: Boldog, Ferenc,
APPLICANT: Casman, Stacie
APPLICANT: Colman, Steve,
APPLICANT: Edinger, Shlomit,
APPLICANT: Gangolli, Baha,
APPLICANT: Kekuda, Ramesh,
APPLICANT: Li, Li,
APPLICANT: Liu, Xiaohong,
APPLICANT: Malyankar, Uriel,
APPLICANT: Miller, Charles,
APPLICANT: Millet, Isabelle,
APPLICANT: Paturajan, Meera,
APPLICANT: Rothenberg, Mark,
APPLICANT: Sciore, Paul,
APPLICANT: Shenoy, Suresh,
APPLICANT: Shimkets, Richard,
APPLICANT: Si, Jingsheng,
APPLICANT: Smithson, Glenda,
APPLICANT: Spytek, Kimberly,
APPLICANT: Stone, David,
APPLICANT: Taupier, Raymond, Jr.,
APPLICANT: Tchernev, Velizar,
APPLICANT: Vernet, Corinne,
APPLICANT: Zernusen, Brian
TITLE OF INVENTION: NOVEL NUCLEIC ACIDS AND POLYPEPTIDES AND METHODS OF USE THEREOF
FILE REFERENCE: 21402-222CIP
CURRENT APPLICATION NUMBER: US/10107, 782
PRIORITY FILING DATE: 2002-03-27
PRIORITY APPLICATION NUMBER: 10/028,248
PRIORITY FILING DATE: 2001-12-19
PRIORITY APPLICATION NUMBER: 60/256, 619
PRIORITY FILING DATE: 2000-12-19
PRIORITY APPLICATION NUMBER: 60/262, 959
PRIORITY FILING DATE: 2001-01-19
PRIORITY APPLICATION NUMBER: 60/272, 408
PRIORITY FILING DATE: 2001-02-28
PRIORITY APPLICATION NUMBER: 60/285, 189
PRIORITY FILING DATE: 2001-04-20
PRIORITY APPLICATION NUMBER: 60/308, 039
PRIORITY FILING DATE: 2001-07-26
PRIORITY APPLICATION NUMBER: 60/311, 266
PRIORITY FILING DATE: 2001-08-09
PRIORITY APPLICATION NUMBER: 60/279, 344
PRIORITY FILING DATE: 2001-03-28

```

```

NUMBER OF SEQ ID NOS: 215
SOFTWARE: Curation version 0.1
SEQ ID NO 35
LENGTH: 7396
TYPE: DNA
ORGANISM: Homo sapiens
FEATURE:
NAME/KEY: CDS
LOCATION: (140) .. (6016)
US-10-107-782-35
Alignment Scores:
Pred. No.: 9,25e-05 Length: 7396
Score: 162.00 Matches: 100
Percent Similarity: 36.34% Conservative: 65
Best Local Similarity: 22.03% Mismatches: 187
Query Match: 8,48% Indels: 103
DB: 12 Gaps: 15

```

US-09-972-758a-2 (1-359) x US-10-107-782-35 (1-7396)

```

QY 2 AlagluProphleuserglutryrln-----HieglProglin--- 14
DB 2107 GCTGATGCTACCTGAGAGACAGACCACTTTGCTCGCTGATCATCTCCACCA 2166
QY 15 -----TherAencyrthnglyAlaAlaValglnglu 26
DB 2167 CGAGAAAGAGCGCGGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 2226
QY 27 -----GluleuapProgluAPro-----Proglu 35
DB 2227 TGTTCGAGGAGCATCGATCTGCGCGAGGCTTCCCAAGAGAGAGAGAGAGAG 2286
QY 36 AlagluAryValProgluAProgluAProgluAProgluAProgluAProgluA 53
DB 2287 GTTTCGAGAGATATGAGATCTGATCTCAACTCCATCTCCAGAGGTTTATGAGCG 2346
QY 54 gluleuglylyarqlyProgluyluglyserleugluserglinProProPro 73
DB 2347 GAAAGAGAGGCGTGGTCCATGATTAAGAGAGAGAGAGAGAGAGAGAGAGAGAG 2406
QY 74 leuglnthrglnAlCyseProgluSerCyS----- 84
DB 2407 TGGCCAGAGCAAAAGCTTCTTCGTCGCGGTGTGTGCGCCCACTGAGAGAGAGAG 2466
QY 85 -----leuargluylgluluglyln 92
DB 2467 CCTGAAGATCACGAGCTCATATAGGTTCCAGGCTGCTGAGAGAGAGAGAGAGAG 2526
QY 93 AenglyAspAspSerSerAlaglylyasphe-----ProProProAlagluVal 109
DB 2527 GAAAGCATTTGCAAGCGCGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 2586
QY 110 GluProthProgluAlagluileuAla-----GlnProCyhisAspserlyAla 127
DB 2587 CGCTCCTACCTGAAGCTGCGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 2646
QY 128 serlysluuglyAlaProAlaAlaglylyglugluutrp-----Glyln 143
DB 2647 GCTGCTGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 2706
QY 144 GlnlnArqln-----leuglylyslvshlsarqlyArqlyProserlysls 160
DB 2707 GGTCAAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 2766
QY 161 lylsArqlysttp-lylProlyrlylysluethrtrpclugluylslvslsPhaspcl 180
DB 2767 TCAGCTCATGCGAGAG-----AAATGAGAGAGAGAGAGAGAGAGAGAGAGAG 2817
QY 180 ulysglinserleuAlaserArqlylearglAluethrleAlaValglngluProVa 200
DB 2818 CGAGCTGTGTGCGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 2877

```







```

Db      2700 CCTGGATGAGAGCCCAACCAAGAGAGATGGCGAGACAAGCTTCGGCGCAGCAA 2759
Qy      336 nleuGlnleuLeuThrGlu---AsnGlnleuHisArgGlnGlnGlu 350
Db      2760 AAAGGCCCTGGAAAGAGATTGTGAACCAAGAAAGACGGGAACAGGAA 2805

```

Search completed: February 5, 2004, 17:11:50  
 Job time : 507 secs

**THIS PAGE BLANK (USPTO)**

GenCore version 5.1.6  
Copyright (c) 1993 - 2004 CompuGen Ltd.

# OM protein - nucleic search, using frame\_plus.p2n model

Run on: February 5, 2004, 13:39:35 ; Search time 94 Seconds  
(without alignments)  
1685.708 Million cell updates/sec

Title: US-09-972-758a-2

Perfect score: 1910  
Sequence: 1 MAEPFLSEYQHPQTSNCTG.....LTENEIHRQOERAPLSPKGD 359

Scoring table:

BLOSUM62	
Xgapop 10.0 , Xgapext 0.5	
Ygapop 10.0 , Ygapext 0.5	
Fgapop 6.0 , Fgapext 7.0	
Delpop 6.0 , Delpext 7.0	

Searched: 569978 segs, 220691566 residues  
Total number of hits satisfying chosen parameters: 1139956

Minimum DB seq length: 0  
Maximum DB seq length: 200000000  
Post-Processing: Minimum Match 0%  
Maximum Match 100%  
Listing first 45 summaries

## Command line parameters:

-MODE=frame.p2n.model -DBV=xlh  
-Q=/cgm2\_1/USPTO.spool/US09972758/runat\_05022004\_095004\_1219/app\_query.fasta\_1.519  
-DB=Issued\_Patents\_NA -QFMT=fastap -SUFFIX=rni -MINMATCH=0.1 -LOOPEL=0  
-LOOPEXT=0 -UNITS=bits -START=1 -END=1 -MATRIX=bloms62 -TRANS=human40.cdi  
-LIST=45 -DOCALLIGN=200 -THR SCORE=pct -THR MAX=100 -THR MIN=0 -ALIGN=15  
-MODE=LOCAL -OUTFMT=pct -NORMext -HEAPSIZE=500 -MINLEN=0 -MAXLEN=200000000  
-USER=US09972758.@CGC\_1.1.56 @runat\_05022004\_095004\_1219 -NCPU=6 -ICPU=3  
-NO\_MMAPP -LARGEQUERY -NEG\_SCORES=0 -WAIT -DSFBILOCK=100 -LONGLOG  
-DBV TIMEOUT=120 -WARN TIMEOUT=30 -THREADS=1 -XGAPOP=10 -XGAPEXT=0.5 -FGAPOP=6  
-FGAPEXT=7 -YGAPOP=10 -YGAPEXT=0.5 -DELOP=6 -DELEXT=7

## Database :

Issued\_Patents\_NA.\*  
1: /cgm2\_6/prodata/1/ina/5A\_COMB.seq.\*  
2: /cgm2\_6/prodata/1/ina/5B\_COMB.seq.\*  
3: /cgm2\_6/prodata/1/ina/6A\_COMB.seq.\*  
4: /cgm2\_6/prodata/1/ina/6B\_COMB.seq.\*  
5: /cgm2\_6/prodata/1/ina/PCUTS\_COMB.seq.\*  
6: /cgm2\_6/prodata/1/ina/backfillseq.seq.\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

## SUMMARIES

Result No.	Score	Query Match	Length DB	ID	Description
1	158.5	8.3	9551	1	US-08-056-200-93
2	158.5	8.3	9551	2	US-08-800-644-93
3	155	8.1	2277	1	US-08-676-967-5
4	155	8.1	2277	2	US-08-676-974-5
5	155	8.1	2277	2	US-09-098-487-5
6	147	7.7	7453	4	US-09-620-312D-248
7	147	7.7	7501	4	US-09-620-312D-249
8	142	7.4	2820	4	US-09-252-991A-2538
9	141	7.4	2304	1	US-08-464-266-1
10	141	7.4	2304	1	US-08-464-272-1
11	141	7.4	2304	3	US-08-464-514-1
12	141	7.4	2304	3	US-08-486-403-1

13	137	7.2	1617	4	US-09-265-013-2	Sequence 2, Appl
14	137	7.2	7308	3	US-09-011-745-3	Sequence 3, Appl
15	137	7.2	7308	3	US-09-011-745-4	Sequence 4, Appl
16	137	7.2	7616	3	US-09-011-745-2	Sequence 2, Appl
17	137	7.2	8202	1	US-08-258-420-13	Sequence 13, Appl
18	137	7.2	8332	3	US-08-850-961-1	Sequence 1, Appl
19	137	7.2	8332	4	US-09-479-776-1	Sequence 1, Appl
20	137	7.2	8332	4	US-09-309-572-11	Sequence 11, Appl
21	137	7.2	8332	4	US-09-315-127-1	Sequence 1, Appl
22	137	7.2	8332	4	US-09-265-013-1	Sequence 1, Appl
23	137	7.2	8332	4	US-09-554-572-25	Sequence 25, Appl
24	134	7.0	3330	3	US-09-162-373-2	Sequence 2, Appl
25	134	7.0	3330	3	US-09-467-946-2	Sequence 2, Appl
26	133.5	7.0	6755	3	US-08-931-999-4	Sequence 4, Appl
27	131.5	6.9	5661	4	US-08-938-105-2	Sequence 2, Appl
28	131	6.9	3023	4	US-09-593-589-10	Sequence 10, Appl
29	131	6.9	6775	4	US-09-620-312D-289	Sequence 80, App
30	130	6.8	1835	4	US-09-216-393B-80	Sequence 28, Appl
31	128.5	6.7	2581	4	US-09-370-838-66	Sequence 66, Appl
32	127	6.6	2139	4	US-09-252-991A-2295	Sequence 2295, Ap
33	126	6.6	7791	3	US-08-949-386-23	Sequence 23, Appl
34	126	6.6	7791	3	US-08-450-562-23	Sequence 23, Appl
35	126	6.6	7791	4	US-08-984-709A-23	Sequence 23, Appl
36	126	6.6	7791	4	US-08-450-272-23	Sequence 23, Appl
37	126	6.6	7808	2	US-08-149-097D-22	Sequence 22, Appl
38	126	6.6	7808	3	US-08-949-386-22	Sequence 22, Appl
39	126	6.6	7808	3	US-08-450-562-22	Sequence 22, Appl
40	126	6.6	7808	4	US-08-984-709A-22	Sequence 22, Appl
41	126	6.6	7808	4	US-08-450-272-22	Sequence 22, Appl
42	125.5	6.6	2384	1	US-07-814-964-10	Sequence 10, Appl
43	125.5	6.6	2384	1	US-08-258-442-10	Sequence 10, Appl
44	125.5	6.6	2384	1	US-08-328-809-5	Sequence 5, Appl
45	125.5	6.6	2384	4	US-08-866-840-5	Sequence 5, Appl

## ALIGNMENTS

RESULT 1  
US-08-056-200-93  
; Sequence 93, Application US/08056200  
; Patent No. 5616500  
; GENERAL INFORMATION:  
; APPLICANT: Steinert, Peter M.  
; APPLICANT: Lee, Seung-Chul  
; APPLICANT: Kim, In-Gyu  
; APPLICANT: Chung, Soo-Il  
; APPLICANT: Park, Sang-Chul  
; TITLE OF INVENTION: Trichohyalin and Transglutaminase-3 and  
; TITLE OF INVENTION: Methods of Using Same  
; NUMBER OF SEQUENCES: 117  
; CORRESPONDENCE ADDRESS:  
; ADDRESSEE: Knobbe, Martens, Olson & Bear  
; STREET: 620 Newport Center Drive, Sixteenth Floor  
; CITY: Newport Beach  
; STATE: CA  
; COUNTRY: U.S.A.  
; ZIP: 92660  
; COMPUTER READABLE FORM:  
; MEDIUM TYPE: Floppy disk  
; COMPUTER: IBM PC compatible  
; OPERATING SYSTEM: PC-DOS/MS-DOS  
; SOFTWARE: PatentIn Release #1.0, Version #1.25  
; CURRENT APPLICATION DATA:  
; APPLICATION NUMBER: US/08/056,200  
; FILING DATE: 30-APR-1993  
; CLASSIFICATION: 435  
; ATTORNEY/AGENT INFORMATION:  
; NAME: Pedrick, Michael F.  
; REGISTRATION NUMBER: 36,799  
; REFERENCE/DOCKET NUMBER: NH054.001A  
; TELECOMMUNICATION INFORMATION:  
; TELEPHONE: (714) 760-0404  
; TELEFAX: (714) 760-9502

```

/ INFORMATION FOR SEQ ID NO: 93:
/   SEQUENCE CHARACTERISTICS:
/       LENGTH: 9551 base pairs
/       TYPE: nucleic acid
/       STRANDEDNESS: single
/       TOPOLOGY: linear
/       MOLECULE TYPE: cDNA
/       HYPOTHETICAL: NO
/       ANTI-SENSE: NO
/   FEATURE:
/       NAME/KEY: CDS
/       LOCATION: 1507..1644
/   FEATURE:
/       NAME/KEY: Intron
/       LOCATION: 1645..2511
/   FEATURE:
/       NAME/KEY: CDS
/       LOCATION: 2512..8070
/   US-08-056-200-93

Alignment Scores:
Pred. No.: 0.002          Length: 9551
Score: 158.50           Matches: 91
Percent Similarity: 39.89%   Conservative: 51
Best Local Similarity: 25.56% Mismatches: 135
Query Match: 8.30%         Indels: 80
DB: 1                      Gaps: 16

US-09-972-758A-2 (1-359) x US-08-056-200-93 (1-9551)

QY 31 GUAATGPRPROGILYALAGLUGUAAG-----VAlProglUGlu 43
   |||||
Db 4810 GAGAGGCGAGCTGAGAGGCGCGAGAGCGCCAGACGCGGAAACAACGGTTTCTCCGGAGAGG 4865
   |||||

QY 44 ApserrATGTGTGInserATGAlaPheProGlnLeuGlyGlyATGPRProglYProglUGly 63
   |||||
Db 4870 GAGAGAAAGAGAGCAGCGC-----GGCGCGCAGAGCAGCGAGAGG 4908
   |||||

QY 64 GUGUySerLeuGluSerGlnProProProLeuGlnThrGlnAlaCyAProglUGuSer 83
   |||||
Db 4909 GAGAAAGAGCTCAGGTTCTCTGAGAGAAAGAGAGCAGCTCCAGCGCGGAGCGCTGCCAA 4968
   |||||

QY 84 CythLeuATrgGluGluGluGly---GlnAnGlyAPaPseSerSerAlaGlyGlyAP 102
   |||||
Db 4969 CAGCTCCAGAGAGAGAGAGCAGCGCTCCAGAG--GGATCAGAGAGAGAGCAGAGAGCA 5027
   |||||

QY 103 PhePro---ProProAlaGluVal-----GluProThPRProglu 114
   |||||
Db 5028 GCGCGCGGACCCAAATAATGAGGTGCGCACTAGAAAGAAAGAAAGAGAGACCGCGCCACAC 5087
   |||||

QY 115 AlAGluLeuLeuAlaGlnProCythLeuAPseSerGluAlaSerLeuGlyAlaProAla 134
   |||||
Db 5088 GCTGTACGCCCAAGCGCAGCCCTAACAAAGAGCAGCTGAGAAAGACA-----GCAAGTCT 5141
   |||||

QY 135 AlAGluGly-GluGluGluUTRGLyGlnGlnGlnAlaArgGlnLeuGlyGlyGlyGlnAla 154
   |||||
Db 5142 GCAGAGAGAGAGAGAGAGCTTACAGAGAGAGAGAGCGCAG-----AA 5183
   |||||

QY 154 GATGATPRProSerLeuGlyGlyGlnAlaTPRLeuProTyTyrGlyGlnThrTPRGLUGl 174
   |||||
Db 5184 GAGAAAGCGCCAAAGAACAGAGACATATCCGCGAGAGAGAGCAGCTGCGACAGAGAGA 5244
   |||||

QY 174 ULyLeuGlyPheAPseGluGlyGlnSerLeuATGAlaSerATGAlaGlnGlnMetP 194
   |||||
Db 5244 AGAGCAGAGCTGTGAGAGAGAAACGGAAGAAAGAAAGAGCAGAGCGGAAAGGCAATTA 5303
   |||||

QY 194 eAlaGlyGluInProValAlaProTyrAsnThrThGlnPheLeuMetAPaPsePheAla 214
   |||||
Db 5304 TCCGAGAGATTAAGAG-----CTGCAGCAGAAAGAGGA 5338
   |||||

QY 214 pGln-----GluGluProAPLeuLeuThrGlyLeuTyrSerLeuValGlnAlaAla 231
   |||||
Db 5337 GCAAGCTGCGGAGAGAGAACCGGAGAGAGAG-----AGCGCGCAGAGAGCG 5381
   |||||

```

QY		231	atysersaspsrThrserrapsrPhmetsGluGluclygllygluglunaprglygl	251
Dd		5382	GGAGAAAAATACCGCAGAGAAAGAGCTTGCACGACGAG-----	54221
QY		251	yserasprglymetcglyyaspqlysergjupeleuGlnarapsPheSerqIurthy	271
Dd		5422	-----GAAGAGCAGCTCTGAGAAGAGAACCGGAGGAAGAG	5456
QY		271	rGIuArg-----tyrhstrngIusertleuGlnametserygl	285
Dd		5457	AAGCCGCGAGACTGGAGAGCGCATGCCCAAAAAGACGACTGCAGCAGGAAGA	5516
QY		285	ngIuLeullelglyutrytreuGluuGluLyCyaleuserargmetcIuaep----	303
Dd		5517	GCACTCTCTAGAGAGAACCGGAGAAABAAGACTCAGAGCGCGAGAGCAATTCG	5576
QY		304	-GluAmaBnarGLuarGLueuGLusertyBarGLueGlyGlyAspApalaArgValAr	323
Dd		5577	GGAGGAAGAGACTGTGACGAGGAAGAGACGACTCTGGAGAGGAACCGGAGACGAG	5636
QY		323	g-----GluLeuGluLeuGluLeu-----AspArgLeuArghlaGluAnle	337
Dd		5637	AAGCGCCAGAGACTGGAGAGCGCAATTCGAAAGAAAGAGAGTGACGACGGAGGA	5696
QY		337	uGlnLeuLeuthrGluuAsn---GluLeuHtARgGInGInGluArg	351
Dd		5697	GCACTCTCTAGAGAGAGAACCGGAGAAAGAACGCCGACGAGAGCGG	5742
RESULT 2				
US-08-800-644-93	/ Sequence 93,	Application US/08800644		
	Patent No.	5958752		
	GENERAL INFORMATION:			
	APPLICANT:	Steinert, Peter M.		
	APPLICANT:	Lee, Seung-Chul		
	APPLICANT:	Kim, In-Gyu		
	APPLICANT:	Chung, Soo-II		
	TITLE OF INVENTION:	Park, Sang-Chul		
	TITLE OF INVENTION:	Trichohyalin and Transglutamlnasee-3 and		
	NUMBER OF SEQUENCES:	117		
	CORRESPONDENCE ADDRESS:			
	ADDRESSEE:	Knobbe, Martens, Olson & Bear		
	STREET:	620 Newport Center Drive, Sixteenth floor		
	CITY:	Newport Beach		
	STATE:	CA		
	COUNTRY:	U.S.A.		
	ZIP:	92660		
	COMPUTER READABLE FORM:			
	MEDIUM TYPE:	Floppy disk		
	COMPUTER:	IBM PC compatible		
	OPERATING SYSTEM:	PC-DOS/MS-DOS		
	SOFTWARE:	Patentin Release #1.0, Version #1.25		
	CURRENT APPLICATION DATA:			
	APPLICATION NUMBER:	US/08/800,644		
	FILING DATE:	14-FEB-1997		
	CLASSIFICATION:	424		
	PRIOR APPLICATION DATA:			
	APPLICATION NUMBER:	US 08/056,200		
	FILING DATE:	30-APR-1993		
	ATTORNEY/AGENT INFORMATION:			
	NAME:	Fedrick, Michael F.		
	REGISTRATION NUMBER:	36,799		
	REFERENCE/DOCKET NUMBER:	NIH054.001A		
	TELECOMMUNICATION INFORMATION:			
	TELEPHONE:	(714) 760-0404		
	TELEFAX:	(714) 760-9502		
	INFORMATION FOR SEQ ID NO:	93:		
	SEQUENCE CHARACTERISTICS:			
	LENGTH:	9551 base pairs		
	TYPE:	nucleic acid		
	STRANDEDNESS:	single		

```

/ TOPOLOGY: linear
/ MOLECULE TYPE: CDNA
/ HYPOTHETICAL: NO
/ ANTI-SENSE: NO
/ FEATURE:
/   NAME/KEY: CDS
/   LOCATION: 1507..1644
/   FEATURE:
/   NAME/KEY: Intron
/   LOCATION: 1645..2511
/   NAME/KEY: CDS
/   LOCATION: 2512..8070
US-08-800-644-93

Alignment Scores:
Pred. No.: 0.002          Length: 9551
Score: 158.50           Matches: 91
Percent Similarity: 39.89% Conservative: 51
Best Local Similarity: 25.56% Mismatches: 135
Query Match: 8.30%      Indels: 80
DB: 2                    Gaps: 16

US-09-972-758a-2 (1-359) x US-08-800-644-93 (1-9551)
QY 31 GUAArgProProGlyAlaGluGluArg-----ValProGluGlu 43
DB 4810 GAGAGGACAGCTGAGGCGCCGAGAGCGCCAGCAGCGGAGAACACGCTTCTCCCGAGAGAG 4869
QY 44 AspSerArgTrpGlnSerArgAlaPheProGlnLeuGlyGlyArgProGluGly 63
DB 4870 GAGAGAGAGAGAGAGCGCC-----GGCCGCGAGAGAGAGAG 4908
QY 64 GluGlySerLeuGlnSerGlnProProProLeuGlnThrGlnAlaCySerProGlnSer 83
DB 4909 GAGAAAGAGCTGACGTTCTCTGAGAGAGAGAGAGAGAGCTCCAGCGCGCGAGCGTCCCAA 4968
QY 84 CysLeuArgGluGlyGlyGlyGlyGlyGlyGlyGlyGlyGlyGlyGlyGlyGlyGlyGlyGly 102
DB 4969 CAGCTCCAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 5027
QY 103 PhePro---ProProAlaGluVal-----GluProThrProGlu 114
DB 5028 GCGCCCGGACCAAAATGAGGTGGCACTAGAAAGAAAGAGAGAGAGAGAGAGAGAGAGAGAG 5087
QY 115 AlaGlnLeuLeuAlaGlnProCySHiAspSerGlnAlaSerLeuGlyAlaProAla 134
DB 5088 GCTGTACGCCAAGCCAGCCCTACAGAGCAGCTGAGGAGAGAGAGAGAGAGAGAGAGAGAG 5141
QY 135 AlaGlyGlyGlyGluGluGluGluGluGluGluGluGluGluGluGluGluGluGluGluGlu 154
DB 5142 GCGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 5183
QY 154 GATGATGProSerLeuLeuLeuLeuLeuLeuLeuLeuLeuLeuLeuLeuLeuLeuLeuLeuLeu 174
DB 5184 GATGATGCGCCAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 5243
QY 174 uLysLysLysPheAspGlnLysSerLeuLeuLeuLeuLeuLeuLeuLeuLeuLeuLeuLeu 194
DB 5244 AGAGCAGCTGCTGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 5303
QY 194 eAlaLysGlyGlnProValAlaProTyrZnThrGlnPheLeuMetAspAspHisAs 214
DB 5304 TCGGAAGATTAAGAG-----CTGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 5336
QY 214 pGln-----GluGluProAspLeuLysThrGlyLeuTyrSerLysArgAlaAlaAla 231
DB 5337 GCAAGCTGCTGGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 5381
QY 231 aLysSerAspAspThrSerAspAspAspPheMetGluGluGlyGlyGluGluAspGlyG 251
DB 5382 GGAGAAAAAATATACCGGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 5421

```

```

QY 251 ySerAspGlyMetGlyGlyAspGlySerGluPheLeuGlnArgAspPheSerGluThrTy 271
DB 5422 -----GAAGAGCAGCTGCTGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 5456
QY 271 rGluArg-----TyrHisThrGlnSerLeuGlnAsnMetSerLysG 285
DB 5457 AAGCGCCAGAGAGTGGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 5516
QY 285 nGluLeuLeuLysGlyTyrLeuGlnLeuLysGlyCysLeuSerArgMetGluAsp----- 303
DB 5517 GCAAGCTGCTGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 5576
QY 304 -GluAsnAsnArgLeuArgLeuGlnSerLysArgLeuGlyGlyAspAspAlaArgValAr 323
DB 5577 GGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 5636
QY 323 g-----GluLeuGluLeuGluLeu-----AspArgLeuArgAlaGluAsnLe 337
DB 5637 AAGCGCCAGAGAGCTGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 5696
QY 337 uGlnLeuLeuThrGluAsn---GluLeuHisArgGlnGlnGlnGlnGlnGlnGlnGlnGln 351
DB 5697 GCAAGCTGCTGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 5742

RESULT 3
US-08-676-967-5
/ Sequence 5, Application US/08676967
/ Patent No. 5747317
/ GENERAL INFORMATION:
/   APPLICANT: COLLINS, KATHLEEN
/   TITLE OF INVENTION: Human Telomerase
/   NUMBER OF SEQUENCES: 10
/ CORRESPONDENCE ADDRESS:
/   ADDRESSEE: Science & Technology Law Group
/   STREET: 268 Bush Street, Suite 3200
/   CITY: San Francisco
/   STATE: CA
/   COUNTRY: USA
/   ZIP: 94104
/ COMPUTER READABLE FORM:
/   MEDIUM TYPE: Floppy disk
/   COMPUTER: IBM PC compatible
/   OPERATING SYSTEM: PC-DOS/MS-DOS
/   SOFTWARE: PatentIn Release #1.0, Version #1.30
/ CURRENT APPLICATION DATA:
/   APPLICATION NUMBER: US/08/676,967
/   FILING DATE:
/   CLASSIFICATION: 530
/   ATTORNEY/AGENT INFORMATION:
/     NAME: Osman Ph.D., Richard A
/     REGISTRATION NUMBER: 36,627
/     REFERENCE/DOCKET NUMBER: UC96-055
/   TELECOMMUNICATION INFORMATION:
/     TELEPHONE: (415)343-4341
/     TELEFAX: (415)343-4342
/   INFORMATION FOR SEO ID NO: 5:
/     SEQUENCE CHARACTERISTICS:
/       LENGTH: 2277 base pairs
/       TYPE: nucleic acid
/       STRANDEDNESS: double
/       TOPOLOGY: linear
/     MOLECULE TYPE: CDNA
US-08-676-967-5

Alignment Scores:
Pred. No.: 0.00055       Length: 2277
Score: 96                Matches: 96
Percent Similarity: 33.93% Conservative: 55
Best Local Similarity: 21.57% Mismatches: 147
Query Match: 8.12%      Indels: 18
DB: 1                    Gaps: 18

US-09-972-758a-2 (1-359) x US-08-676-967-5 (1-2277)

```

Qy	6	LeuSerGluTyrGlnHisAlaGlnProGlnThrSerAsnGlySerHisValAlaAlaValAlaGln	25
Db	144	CTAGGTGACCTTCACAGAT-----GCTGGAGAGACGTGCACGGCGC	182
Qy	26	GluGlu-----LeuAsnProGluArgProProGluValAlaGluGlu	38
Db	183	CCTGAAGAGATACACACCTTCGAGGGCTGCAGAAATCAACGTGACCTGGCCAGAAAGA	242
Qy	39	ArgValProGluGluAspSerArgTrpGlnSerArg-----AlaPheProGlnLeuGly	56
Db	243	GCTCGCAACAAGACCAAGAGAAAGGGAAGAACAGAAACAGCGAGTCCCCCAAGAAAGA	302
Qy	57	GlyArgProGlyProGluGlyGluGlySerLeuGlnSerGlnPro-----	71
Db	303	GGCCCA---GGCCAGAGAGCCAGAGAGTGGCCGACAGAAAGCCCGCTGATCATCCGCA	359
Qy	72	ProProLeuGlnThrGlnAlaCysPheProGluSerSerCysLeuArgGluGlyGluIleGly	91
Db	360	CTTAGAGCTTCAAGTGCACGAGACGACCTTGAAAGCCGTGTTCCGCCAGTTCCGGGC---	416
Qy	92	GlnAsnGlyAspAspSerSerAlaGlyGlyAspPheProProProAlaGluValGluPro	111
Db	417	-----CGTGTCTGGAAGTGAACATCCCCCGCAAGCCCGCAAGCGCA---	455
Qy	112	ThrProGlnAlaGluLeuLeuAlaGlnProCysHisAspSerGluAlaSerTyrLeuGly	131
Db	456	-----GATGGCGCGCTTCGGCTTCGACTTCAGTAACCTGCTGAGGCCGGGCAAGGC	509
Qy	132	AlaProAlaAlaGlyGlyGluGluGlu-----	140
Db	510	-----CCTGAAGGCGATGACATGAAGAAGATCAAGGGCCGCAACCGGCGGTGAGCTG	563
Qy	141	-----TrpGlyGlnGlnGlnArgGlnLeuGlyIleValValValHisArgArg-ArgPr	157
Db	564	GGCGGTGGCCAAAGACCAAGTACAAAGACACCCAGACGCTGAGCGCATCGCGAGAGAA	623
Qy	157	AsnTyrValValValValGlnHisTrpTyrProTyrTyrTyrValLeuThrTrpGluGluValVal	177
Db	624	GAGCCACGAGCAAGCAACGACGAGAGGCTGAAAGAAAGGCGCGAGAGAGAGACAT	683
Qy	177	AspAspGluValValGlnSerLeuArgAlaSerArgIleArgAlaGluMetPheAlaValGlu	197
Db	684	GGAGAGAGAGAGAAACGAC-----	702
Qy	197	GlnProValAlaProTyrAsnThrThrGlnPheLeuMetAspAsnHisAspGlnGluG	217
Db	703	-GAGCAGCAGACGACGACGACGACGAGAGAGACGCGCTGTTCCAGCAGAGACGAGAGGA	761
Qy	217	uProAspLeuValThrGlyLeu-----TyrSerTyrValArgAlaAla-----	230
Db	762	GGAGAACATTCAGAGCAAGGCTGACCAAGCCCGTGCAGATCCAGAAAGCCGCGGTGAAAG	821
Qy	231	-----AlaValSerAspAspThrSerAspAspAspPhe---MetGluGluGlyG	246
Db	822	CCCCGGCCCCCGCAAGACGACGACGACGACGAGGAGGAGGACGCGACCTGGAGAGAGGGA	881
Qy	246	GlyGluValAspGlyGlySerAspGlyMetGlyValAspGlySerGluPheLeuGlnArgAs	266
Db	882	CAGCATGCAC-----GACCGCAGAGAGCTGGCCCAAGGGA	917
Qy	266	PheSerGluThrTyrGluArgTyrHisThrGluSerLeuGlnAsnMetSerTyrGln--	285
Db	918	CACCAAGACCGAGAGCAGAGACAAAGGCGCTGCAGGTGAGCAACAAAGAAAGAACGCA	977
Qy	285	-----	285
Db	978	GCTGCCACGCACTGAACGAGGCGAAGACCGTGTTCATCCGCAACCTGACCTTGACAG	1037
Qy	286	-----GluLeuIleValGluTyrLeuGluLeu-----	294
Db	1038	CGAGAGGAGAGAGCTGGGCGACCTCTCTGCAGAGATGTTCCGACAGCTGAAGATACGTCCGAT	1097

Qy	294	-----	294
Db	1098	CGTCTGCACCCCGACACCGACACAGCAAGGCTGGCCCTTCCGCCAGTTCATGACCCA	1155
Qy	295	-----GluIyScyAeu-----SerArgMetGluAspGluAsnAsnArgLeuAr	309
Db	1158	GGAGCGCCCGCAGAGTGTCTGTGGCCGCGCCAGCCCGAGAGACGAGCCCGGCGCTGAA	1217
Qy	309	gIeuGluSerLeuArgLeuGlyGlyAspAlaArgValArgGluLeuGluLeuGluLe	329
Db	1218	GCTGGAGCGCCGCGCAGCTG-----AAGTGGACCTGGCCGT	1253
Qy	329	uAspArgLeuArgAlaGluAsnLeuGlnLeuThrGluAsnGluLeuHisArgGlnG1	349
Db	1254	GACCCCGACGAGGCGCCCGCAAGCTGCAGACCACCAAGGT-GAAGAACCCACCGGACCC	1312
Qy	349	ngIuArgAlaPro 353	
Db	1313	GCAACCTGTACT 1325	
RESULT 4			
US-08-676-974-5			
Sequence 5, Application US/08676974			
Patent No. 5770422			
GENERAL INFORMATION:			
APPLICANT: COLLINS, KATHLEEN			
TITLE OF INVENTION: Human telomerase			
NUMBER OF SEQUENCES: 10			
CORRESPONDENCE ADDRESS:			
ADDRESSEE: Science & Technology Law Group			
STREET: 268 Bush Street, Suite 3200			
CITY: San Francisco			
STATE: CA			
COUNTRY: USA			
ZIP: 94104			
COMPUTER READABLE FORM:			
MEDIUM TYPE: Floppy disk			
COMPUTER: IBM PC compatible			
OPERATING SYSTEM: PC-DOS/MS-DOS			
SOFTWARE: PatentIn Release #1.0, Version #1.30			
CURRENT APPLICATION DATA:			
APPLICATION NUMBER: US/08/676,974			
FILING DATE:			
CLASSIFICATION: 530			
ATTORNEY/AGENT INFORMATION:			
NAME: Osman Ph.D., Richard A			
REGISTRATION NUMBER: 36,627			
REFERENCE/DOCKET NUMBER: UCB96-055			
TELECOMMUNICATION INFORMATION:			
TELEPHONE: (415)343-4341			
TELEFAX: (415)343-4342			
INFORMATION FOR SEQ ID NO: 5:			
SEQUENCE CHARACTERISTICS:			
LENGTH: 2277 base paire			
TYPE: nucleic acid			
STRANDEDNESS: double			
TOPOLOGY: linear			
MOLECULE TYPE: CDNA			
US-08-676-974-5			
Alignment Scores:			
Pred. No.:	0.00055	Length:	2277
Score:	155.00	Matches:	96
Percent Similarity:	33.93%	Conservative:	55
Best local Similarity:	21.57%	Mismatches:	147
Query Match:	8.12%	Indels:	148
DB:	1	Gaps:	18
US-09-972-758A-2 (1-359) x US-08-676-974-5 (1-2277)			
Qy	6	leuSerGluYrGlnHnSGlnProGlnThrSerAsnCystrHnGlyAlaAlaValGln	25
Db	144	CTAGTGACCTTCAGCAT-----GCTGAGAGACGTGACCGCGC	182



```

QY 26 GIUGLU-----LeuAsnProGluArgProProGluValaGluGlu 38
DB 183 CCTGAAGAGATCACCACTTGAGAGGCTGCAAGATCAAGTGCAGCGCCAGAGAA 242
QY 39 ArgValProGluGluAspSerArgTrpGlnSerArg-----AlaPheProGlnLeuGly 56
DB 243 GCTGGCCACCAAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 302
QY 57 GluArgProGluProGluGluGluGluGluGluGluGluGluGluGluGluGlu 71
DB 303 GCCCAA---GGCCAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 359
QY 72 ProProLeuGlnThrGlnAlaCysProGluSerSerCysLeuArgGluGluGluGlu 91
DB 360 CCGTAGCTTCAAGTGCAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 416
QY 92 GlnAsnGlyAspAspSerSerAlaGlyGlyAspPheProProAlaGluValGluPro 111
DB 417 -----CGTGTGAGAGGTGAACATCCCGCCAGAGCCGAGCGCA--- 455
QY 112 ThrProGluAlaGluLeuLeuAlaGlnProCysHisAspSerGluAlaSerLysLeuGly 131
DB 456 -----GATGCGCGGCTTGGCTTGGTGCAGTTCAAGAACCTGCTGAGAGCGGCAAGGC 509
QY 132 AlaProAlaAlaGlyGluGluGluGlu----- 140
DB 510 -----CCTGAAGGAGATGACATGAAGAGAGATCAAGGCGCCGAGCGGTGAGCTG 563
QY 141 -----TrpGlyGlnGlnGlnArgGlnLeuGlyLysLysLysHisArgArg-ArgPr 157
DB 564 GGCGGTGCGCAAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 623
QY 157 SerLysLysLysArgHisTrpLysProTrpTrpLysLeuThrTrpGluGluLysLys 177
DB 624 GAGCCACGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 683
QY 177 sPheAspGluLysGlnSerLeuArgAlaSerArgAlaGluMetPheAlaLysGlu 197
DB 684 GAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 702
QY 197 yGlnProValAlaProTrpAsnThrThrGlnPheLeuMetAspAspHisAspGlnGlu 217
DB 703 -GACGACGACGACGACGACGACGACGACGACGACGACGACGACGACGACGACGACG 761
QY 217 uProAspLeuLysThrGlyLeu-----TyrSerLysArgAlaAla----- 230
DB 762 GGAGAACATCGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 821
QY 231 -----AlaLysSerAspAspThrSerAspAspAspPhe---MetGluGluGluGlu 246
DB 822 CCCCCCCCCCGCCAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 881
QY 246 yGluGluAspGlyGlySerAspGlyMetGlyGlyAspGlySerGluPheLeuGlnArgAs 266
DB 882 CAGCATCGAC-----GACGGCGAGAGAGCTGGCCAGAGCGA 917
QY 266 pPheSerGluThrTrpGluArgTrpHisThrGluSerLeuGlnAspMetSerLysGln-- 285
DB 918 CACACACACACGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 977
QY 285 ----- 285
DB 978 GCTGCCAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1037
QY 286 -----GluLeuLysGluGluGluGlu----- 294
DB 1038 CGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1097
QY 294 ----- 294
DB 1098 CGTGTGACACCCGAGACACGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1157

```

```

QY 295 -----GluLysCysLeu-----SerArgMetCysAspGluAsnAsnArgLeuAr 309
DB 1158 GGAGGCCCGCCAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1217
QY 309 GluGluLysSerLysArgLeuGluGlyGlyAspAspAlaArgValArgGluLeuGluGlu 329
DB 1218 GCTGAGAGCGCGCCAGCTG-----AAGGTGAGACTTGCGCGT 1253
QY 329 uAspArgLeuArgAlaGluAsnLeuGlnLeuLeuThrGluAsnGluLeuHisArgGlnGlu 349
DB 1254 GACCCGCGACGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1312
QY 349 nGluArgAlaPro 353
DB 1313 GCACCTGTACTT 1325

RESULT 5
US-09-098-487-5
/ Sequence 5, Application US/09098487
/ Patent No. 5917025
/ GENERAL INFORMATION:
/ APPLICANT: COLLINS, Kathleen
/ TITLE OF INVENTION: Human Telomerase
/ NUMBER OF SEQUENCES: 11
/ CORRESPONDENCE ADDRESS:
/ ADDRESSEE: Science & Technology Law Group
/ STREET: 268 Bush Street, Suite 3200
/ CITY: San Francisco
/ STATE: CA
/ COUNTRY: USA
/ ZIP: 94104
/ COMPUTER READABLE FORM:
/ MEDIUM TYPE: Floppy disk
/ OPERATING SYSTEM: IBM PC compatible
/ SOFTWARE: PatentIn Release #1.0, Version #1.30
/ CURRENT APPLICATION DATA:
/ APPLICATION NUMBER: US/09/098,487
/ FILING DATE:
/ CLASSIFICATION:
/ ATTORNEY/AGENT INFORMATION:
/ NAME: Osman Ph.D., Richard A
/ REGISTRATION NUMBER: 36,627
/ REFERENCE/DOCKET NUMBER: UCB96-055
/ TELECOMMUNICATION INFORMATION:
/ TELEPHONE: (415)343-4341
/ TELEFAX: (415)343-4342
/ INFORMATION FOR SEQ ID NO: 5:
/ SEQUENCE CHARACTERISTICS:
/ LENGTH: 2277 base pairs
/ TYPE: nucleic acid
/ STRANDEDNESS: double
/ TOPOLOGY: linear
/ MOLECULE TYPE: cDNA
/ US-09-098-487-5

Alignment Scores:
Pred. No.: 0.00055 Length: 2277
Score: 155.00 Matches: 96
Percent Similarity: 33.93% Conservative: 55
Best Local Similarity: 21.57% Mismatches: 147
Query Match: 8.12% Indels: 148
DB: 2 Gaps: 18

US-09-972-758a-2 (1-359) x US-09-098-487-5 (1-2277)
QY 6 LeuSerGluTrpGlnHisGlnProGlnThrSerAsnCysThrGlyAlaAlaValGln 25
DB 144 CTACGTGACCTTCAGCAT-----GCTGGAGAGAGCTGACAGCGCGC 182
QY 26 GIUGLU-----LeuAsnProGluArgProProGluValaGluGlu 38
DB 183 CCTGAAGAGATCACCACTTGAGAGGCTGCAAGATCAAGTGCAGCGCCAGAGAA 242

```

Oy		39	ArgValProGluGluAspSerAryTrpGlnSerArg-----AlapheProGluLeuGly	56
Db		243	GCTCGGCACAGACCAGAGAAAGGGCCAGAACAAGAACGCCAGTCCCCCAAGAAAGA	302
Oy		57	GlyArpProGlyProGluGlyGluGlySerLeuGluSerGlnPro-----	71
Db		303	GCCCAA---GGCCAAGAAAGCCCAAGGTGGCCGACAGAAAGGCCCGCCTGATCATCCGGA	359
Oy		72	ProProlenGlnThrGlnAlaCyseProGluSerSerCybleuAryGluGlyGlySerGly	91
Db		360	CCTAGAGCTTCAGATGCACGCGAGACGACTGTAAGACCGTGTTCCGCCAGTTCCGAGC---	416
Oy		92	GlnAnenGlyAspaBspSerSerAlaGlyVaryAspHeProProProlaGluValGluPro	111
Db		417	-----CGTGCTGGAGATAACATCCCCTGGCAAGCCCGCAAGCGGAAA---:	455
Oy		112	ThrProGluAlaGluLeuLeuAlaGlnProCyshAspSerGluAlaSerIyleuGly	131
Db		456	-----GATGGCGGGCTTCGGCTTCCTGTCACATTCAAGAACTGCTGAGAGCGGGCAAGC	509
Oy		132	AlaProlaAlaGlyGlyGluGluGln-----	140
Db		510	-----CCTGAAGGGCATGAACATGAAGAGATCAAGGGCCGCAACCGTGCCGTGACTG	563
Oy		141	-----TrpGlyGlnGlnGlnArgGlnLeuGlySylSylSylShisAryArg-AryPr	157
Db		564	GGCCGTGGCCCAAGGACAAAGTACAAAGACACCCAGACCTTGAGCCGCATCGCCAGAGAGA	623
Oy		157	oSerIyleSylSylAryShisTrpLysProTyTrTyIyleuThrTrpGluGlySylSyl	177
Db		624	GAGCCACGAGACGACGACCCAGAGAGGCTGAAGAAGAGGGCCGCGAGAGAGAGCAT	683
Oy		177	sPheAspGlyIsgInserIleuAryAlaSerAryGlleaArgAlaGluMetPheaIalysGI	197
Db		684	GGAGAGAGAGAGAAACGAC-----	702
Oy		197	yGlnProValAlaProTyTrpAntThrThrGlnPheLeuMetAspAryhisAryGlnGlu	217
Db		703	-GAGCACGACGACGACGACGACGAGAGAGACGGCGTGTTCAGACGACGAGACGAGAGAGA	761
Oy		217	uProAspLeuIyleThrGlyLeu-----TySerIylSAryAlaIala----	230
Db		762	GGAAhAAtCTCAGAGACGAAAGTAGTACCAAGACCCTGACATCCAGAAAGCCGCCGTGAAGC	821
Oy		231	-----AlaIylSerAspAryThrSerAspAspAryhe--MetGluGluGly	246
Db		822	CCCCGCCCCGCCAAGACGACGACGACCAACAGAGAGAGACGACACTGAGAGAGAGCGA	881
Oy		246	yGluGluAspGlyGlySerAspGlyMetGlyIlyAspGlySerGluPheLeuGlnAryAs	266
Db		882	CAGCATTCAC-----GACCGCAGAGAGCTGGCCCAAGAGGA	917
Oy		266	pHeserGlnThrIyrGlyAryGlyTrpHisThrGluSerLeuGlnAspMetSerIylGln-	285
Db		918	CACCAgAcCCAgAGAcAGAGAGACAGACAGCCGTGCAGGTGAGCAACAAGAAAGCGCAA	977
Oy		285	-----	285
Db		978	GCTGCCACGACGtGMAAGAGAGGCAACCGTGTTCATCCGCAACCTGACCTTCAGACAG	1037
Oy		286	-----GluLeuIleLyGluGlyTrpLeuGluLeu-	294
Db		1038	CGAGAGAGAGAGAGCTGGCGAGCTGCTGCACAGATTCCGCGAGCTGAAGTACGTGCGAT	1097
Oy		294	-----	294
Db		1098	CGTGTGCACCCCGACACCGAGACACAGCAAGAGGCTGCCTTCGCCCACTTCATAGCCCA	1157
Oy		295	-----GluIylScYblau-----SerAryMetGlnAspGluAsnAspAryLeuAr	309
Db		1158	GGAGGCGCCGCCAGAAAGTCTCTGTCGGCCGACAGCCCGCCAGAAACAGAGCGCGCGCTTADA	1217

```

QY      309 gIeGluSerIyArGluEnGIyGLyAsPaAPAlArGVAlArGluLeuGIuLeuGIuLe 329
          |||::: |:::||||| ::::::::::||| ::
Db      1218 GCTGGAGCGGCCCGCACTG-----AAGTGCACCTGGCCCT 1253

QY      329 uASPArGLeArGLAgluSnleuGInleuThrGlusnGluLeuHIsarGInGI 349
          |||::: |:::||||| |||::: |||
Db      1254 GACCCTGAGCAGAGCGCCGCAAGCTGCAGACACCAAGGT-GMAGAAGCCCGCCGACACC 1312

QY      349 nGUArGLAPro 353
          ::: |||
Db      1313 GCMAACTGTACT 1325


RESULT 6
US-09-620-312D-248
; Sequence 248, Application US/09620312D
; Patent No. 6569662
; GENERAL INFORMATION:
; APPLICANT: Tang, Y. Tom
; APPLICANT: Liu, Chenghua
; APPLICANT: Asundi, Vinod
; APPLICANT: Zhang, Jie
; APPLICANT: Ren, Feiyun
; APPLICANT: Chen, Rui-hong
; APPLICANT: Zhao, Qing A.
; APPLICANT: Wehrman, Tom
; APPLICANT: Xue, Aidong J.
; APPLICANT: Yang, Yonghong
; APPLICANT: Wang, Jian-Rui
; APPLICANT: Zhou, Ping
; APPLICANT: Ma, Yunqing
; APPLICANT: Wang, Dunhui
; APPLICANT: Wang, Zhiwei
; APPLICANT: John Tillinghast
; APPLICANT: Drmanac, Radcoje T.
; TITLE OF INVENTION: No. 6569662e1 Nucleic Acids and
; FILE REFERENCE: Polypeptides
; FILE REFERENCE: 784CIPB2B
CURRENT APPLICATION NUMBER: US/09/620,312D
CURRENT FILING DATE: 2000-07-19
PRIOR APPLICATION NUMBER: 09/552,317
PRIOR FILING DATE: 2000-04-25
PRIOR APPLICATION NUMBER: 09/488,725
PRIOR FILING DATE: 2000-01-21
NUMBER OF SEQ ID NOS: 1105
SOFTWARE: pc_fl_genes Version 1.0
SEQ ID NO 248
; LENGTH: 7453
; TYPE: DNA
; ORGANISM: Homo sapiens
FEATURE:
; NAME/KEY: CDS
; LOCATION: (1)..(4362)
US-09-620-312D-248


Alignment Scores:
Pred. No.: 0.0136 Length: 7453
Score: 147.00 Matches: 103
Percent Similarity: 36.84% Conservative: 51
Best Local Similarity: 24.64% Mismatches: 170
Query Match: 7.70% Indels: 95
DB: 4 Gaps: 15


US-09-972-758A-2 (1-359) x US-09-620-312D-248 (1-7453)
QY      15 ThSeRanCySThrGlyAlaAlaValGlnGlu-----GIuLeuAsnProGIuAr 32
          |||::: |:::||||| ::::::::::||| ::
Db      3948 ACTACAGATCACACAGAGCTTTCTTCAGAGGAGACACAGCCAGAACTTAACCTTGAGCAG 4007

QY      33 ProBGrluAlaGIuGIuArGVAlProGIuGluSpsErArGPrgInSerArGLaPhe 52
          |||::: |:::||||| |||::: |||
Db      4008 TCGGATCCGCGAGCTGGGAAGAGGAAGAACAAGTCTTCA----- 4046

QY      53 ProGInleuGIyGLyArGVProGIyProGIuGIuGIuSerIeuGIuSerGIuProPro 72

```

[illegible]

```

Db      5096 TCTGAGGAGACCCGCGACACGCGCAGACGAGAGAGATAGCTGGCG 5143

RESULT 7
US-09-620-312D-249
: Sequence 249, Application US/09620312D
: Patent No. 6569662
: GENERAL INFORMATION:
: APPLICANT: Tang, Y. Tom
: APPLICANT: Liu, Chenguang
: APPLICANT: Asundi, Vinod
: APPLICANT: Zhang, Jie
: APPLICANT: Ren, Feiyun
: APPLICANT: Chen, Rui-hong
: APPLICANT: Zhao, Qing A.
: APPLICANT: Wehrman, Tom
: APPLICANT: Xue, Aidong J.
: APPLICANT: Yang, Yonghong
: APPLICANT: Wang, Jian-Rui
: APPLICANT: Zhou, Ping
: APPLICANT: Ma, Yungqing
: APPLICANT: Wang, Dunrui
: APPLICANT: Wang, Zhiwei
: APPLICANT: John Tillinghast
: APPLICANT: Dmanac, Radoje T.
: TITLE OF INVENTION: No. 6569662el Nucleic Acids and
: FILE OF INVENTION: Polypeptides
: FILE REFERENCE: 784CIP2B
: CURRENT APPLICATION NUMBER: US/09/620,312D
: PRIOR FILING DATE: 2000-07-19
: PRIOR APPLICATION NUMBER: 09/552,317
: PRIOR FILING DATE: 2000-04-25
: PRIOR APPLICATION NUMBER: 09/488,725
: PRIOR FILING DATE: 2000-01-21
: NUMBER OF SEQ ID NOS: 1105
: SOFTWARE: pt_FL_genes Version 1.0
: SEQ ID NO 249
: LENGTH: 7501
: TYPE: DNA
: ORGANISM: Homo sapiens
: FEATURE:
: NAME/KEY: CDS
: LOCATION: (1)..(4410)
US-09-620-312D-249

Alignment Scores:
Pred. No.: 0.0137 Length: 7501
Score: 147.00 Matches: 103
Percent Similarity: 36.84% Conservative: 51
Best Local Similarity: 24.64% Mismatches: 170
Query Match: 7.70% Indels: 95
DB: Gaps: 15

US-09-972-758A-2 (1-359) x US-09-620-312D-249 (1-7501)
QY 15 Thiserascysthrgllylaalaalavalglnlu-----gluleuasnprogluarg 32
Db 3996 ACTACAGATACACAGAGGCTTCTTCAGAGAGACACGCCAGAACTAAACCTGAGCAG 4055
QY 33 ProerolglylaiglualuargvalProglugluNapserargTpglnuserAylahe 52
Db 4056 TCGGATCCGGAGCTGGAGAGGAGGAAGAAACAGCTTCA----- 409
QY 53 ProglinleuglylglyargProglyProgluglylgluglyserleuglnuserGlnProBo 72
Db 4095 GAGACACAGAGAGAGAGAGAGAGCGCAGAGAAACCTGGAGAAAGCAAGTGTGGCCCT 415
QY 73 ProleuglnThrglnalAcySProgluserSerCyseuargglugly-----Glu 89
Db 4155 GCAGTCCACAGTGGCTGATACCAAGAAAGAAAGTGAATGATGACACCTGGAAACAAATTGAAG 4214
QY 90 LynglyglinleuglylaapaspserSerAlagllylaspPheProProAlaglual 109
Db 4215 TCTGGAGAAAGCAAGAAAGGCTTCTGAAAGAGCGCGAGGCCCTTGAGCCAGCG----- 4266

```

QY 110 GlnProthrProgluAgluLeuAgluProCysHisAspSerGluAAspLeu 129  
Db 4269 ---CCTGAGAGAGAGGCACTGGCGTATGACAAACTGGAGAAAGCAAGAACGGCTGCA 4325  
QY 130 LeuGluAAsp-----ProAlaAglGlyGluGluTrp 141  
Db 4326 GCAGGACCTGGACGACCTCAACGGTGAACCTGGACCAACAGCGGACGGCTCAACT 4385  
QY 142 GlyGlnGlnGlnGln----- 147  
Db 4386 GGAAGAAAGACAGAAAGATTGTGACAGCTGTAGCAGAGAAAGACATCTGCTC 4445  
QY 148 -----LeuGlyLysLysHisArgArgProSerLysLys 161  
Db 4446 GCTATGCCGAAGACGGGACCGGGCCGAAGCCGAGGCCAGAGAAAGAAACCAAGCCC 4505  
QY 162 ArgHisTrpLysProTyrTyrLysLeuThrTrpGluGluLysLysPheAsp---Glu 180  
Db 4506 TGTCACTGCGCCCGGCGCTCGAGGAAGCCCTGGAGGCC-AAAGAGAGATTGAAGGCGAG 4564  
QY 181 LysGlnSerLeuArgAAsp----- 187  
Db 4565 AACAAAGAGCTCCGAGCAGACATGAAGACCTCATGAGCTCAAGATGATGTGGAAAA 4624  
QY 188 -----ArgIleArgAlaGluMetPheAlaLysGlnProValAlaProTyrAsn 204  
Db 4625 AACGTTCAAGAACTTGAAAAATCCAAAGGGCGCTTAGAGCGAGGTGGAGAAATGAG 4684  
QY 205 ThrThrGlnPheLeuMetAspAspHisAspGlnGluTrpAspLeuLysThrGlyLeu 224  
Db 4685 ACCCAAGCTGGAGAGCTGGAAAGCAAACTCAGAGCCAGGAAGATGCCAAGCTTCGCTG 4744  
QY 225 TyrSerLysArgAlaAlaAlaLysSerAsp-----AspThrSerAspAsp 239  
Db 4745 GAGGTCAACATGACAGCGCATGAAAGCGGAGTTGCAGAGAGACGTCGAAACAGAGGATGAG 4804  
QY 240 --AspPheMetGluGluGlyGlyGluGluAspGlyGlySerAspGlyMetGlyLysArg 259  
Db 4805 CAGAAATGAAGAAAGAAAGCGGCTGCTGATCAAAAGAGTGGGAGCTCGAGCGGAGCTG 4864  
QY 259 LysSerGluPheLeuGlnArgAspPheSerGlnThrTyrGluArgTyrHisThrGluSer 278  
Db 4865 GAGATGAGAGAAACACGGGGCGCTGCTTACTTCAAGAAAGAAAGATGAGATGAC 4924  
QY 279 LeuGln-----AsnMetSerLysGlnLeuLeuLys 289  
Db 4925 CTGAAGAGCCTCGAAGCCCAAAATGAGAGCTCGAACAAGCTCGGATGAGGTATTAAg 4984  
QY 290 GluTyrLeuGluLeuGluLysCysLeuSerArgMetGluAspGluAsnAsnArgLeuArg 309  
Db 4985 -----CAGCTCCGCAAGCTCCAGGCTCAGATGAAGATTTCACCAAGT-----GAA 5029  
QY 310 LeuGluSerLysArgLeuGlyLysAsp-----AlaArgValArgGluLeuGluLeu 327  
Db 5030 TTAAAGAAAGCTGTGCATCCAGAGATGAGATTTTGCTCATCAACAAAGAGAGTGAAGAAg 5089  
QY 328 GluLeuAspArgLeuArgAlaGluAsnLeuGlnLeuLeuThrGluAsnGluLeu----- 345  
Db 5090 AAATTGAAGACTCTGGAAGCAAGAAATCTTCAATTG-----CAGAGAGAACTTGCTCA 5143  
QY 346 -----HisArgGlnGlnGluArgAlaProLeuSer 355  
Db 5144 TCTGAGCGAGCGCGCCGACAGCGGAGCAGAGAGATGAGTGGCGG 5191

RESULT 8  
US-09-252-991A-2538  
; Sequence 2538, Application US/09252991A  
; Patent No. 6551795  
; GENERAL INFORMATION:  
APPLICANT: Marc J. Rubenfield et al.  
TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO PSEUDOMONAS  
AERUGINOSA FOR DIAGNOSTICS AND THERAPEUTICS

```

/ FILE REFERENCE: 107196.136
/ CURRENT APPLICATION NUMBER: US/09/252,991A
/ CURRENT FILING DATE: 1999-02-18
/ PRIOR APPLICATION NUMBER: US 60/074,788
/ PRIOR FILING DATE: 1998-02-18
/ PRIOR APPLICATION NUMBER: US 60/094,190
/ PRIOR FILING DATE: 1998-07-27
/ NUMBER OF SEQ ID NOS: 33142
/ SEQ ID NO 2538
/ LENGTH: 2820
/ TYPE: DNA
/ ORGANISM: Pseudomonas aeruginosa
/ US-09-252-991A-2538

Alignment Scores:
Pred. No.: 0.00946 Length: 2820
Score: 142.00 Matches: 98
Percent Similarity: 31.26% Conservative: 48
Best Local Similarity: 20.99% Mismatches: 156
Query Match: 7.43% Indels: 165
DB: 4 Gaps: 19

US-09-972-758A-2 (1-359) x US-09-252-991A-2538 (1-2820)
QY 12 GlnProGlnThrSerAsnGlySerThrGlyAlaAlaAlaValGlnGluGluLeuAsnProGlu 31
Db 139 CAACCTGAAGCGGAGTCCGGCAACTCT-----CGCCGACATACGACCTCGT 183
QY 32 ArgProGluGlyAlaGluGluArgValProGluGluAspSerArgTrpGlnSerArgAla 51
Db 184 CGCCCTCCAGGAAGCGGACGCGGCGAGCTCGTTCCGGCTACGTCATCAGGTCCGAGCA 243
QY 52 PheProGlnLeuGlyGly-----ArgProGlyProGluGly 63
Db 244 CTGGGCCCGTTGGGGGGCGTTCCCTACTGStrACAGAGCTCAACCGCAACTCGGGCG 303
QY 64 GluGlySerLeuGlnSerGlnPro---ProProLeuGlnThrGlnAlaCysProGluSer 82
Db 304 TATCGCCCAAGCAGACAGCAAGGCTCTGACGCCCTGGCGGCGGACCTCGTGAAGATCA 353
QY 83 SerCysLeuArgGluGly----- 88
Db 364 TCCGTTGGCCCGGATCCGCGCGGAGCGGGCGCATCTTCGTGGCGTTTGGCAAGCGCGCA 423
QY 89 GluLysGlyGlnAsnGlyAspAspSerSerAlaGlyLysPhePro---ProProAla 107
Db 424 CGCGCTGGTGAATGATGATGATCATCTGCTCGCGGCTCGACCGCACCGCGAGCT 483
QY 108 GluValGluProThrProGluAlaGluLeu----- 117
Db 484 CGCCTATATCCGAGCTGATCGGCACTACCGCCATCAGTGTGATGGGCAACATGAA 543
QY 118 -----LeuAlaGlnProCysHisAspSerGlu 126
Db 544 CACCCACGCGCGTGCATCTTCGGAATACTGCGCCCTGGCGGACCTCGGCGTGAATGGGCC 603
QY 127 Ala-----SerLysLeuGlyAlaProAla----- 134
Db 604 GCAGGTGAGGAGCGACGTTTCCACGCTGGCGGCGGCGGACGCAATCTGATCACAATCTGCT 663
QY 134 ----- 134
Db 664 CAGTTCGATCTCGCTTCGAGCGGGTGCAGCTGCTGCACGACCGCATTCGACCATCT 723
QY 135 -----AlaGlyGlyGluGluGluTrpGlyGlnGlnGlnArgGln---LeuGlyLysLys 151
Db 724 GCCGGATACGGTGGAGATCTCGCTCCGGAACAACCTGCGGACAAAGGCTTTGGTGGCG 783
QY 152 LysHisArgArgArgProSerLysLysLys-----ArgHisTrpLysProTyrTyr 168
Db 784 CGGCGATGCGCGGAGATCGCGCATGATGTCGCGACGATCCAGCGCTGGAAGAACATAC 843
QY 169 LysLeuThrTrpGluGluLysLysLysPheAspGlnLysGlnSerLeuAlaGlnSerArg 188

```

```

Db      844 CTGGAGACATTCAGACGAGGCGCTTGCAACGGCGCTGGGATCCCGGATCGACTG 903
Qy      189 TLeaGAlaGluMeCHeaLalYsGlyGlnProValAlaProTyrAnThrGlnPhe 208
Db      904 CTGCGG----- 909
Qy      209 LeuMeCAspAspHisAspGlnGlnGluProAspLeuLysThrGlyLeuTyrSerLysArg 228
Db      910 -----CGCGGCTCTGGGCTGACAGT 930
Qy      229 AlaAlaAlaLysSerAspAspThrSerAspAspPheMetGluLuu----- 244
Db      931 CTGGCGCGCGGACGAGCGAGCAAGGCGTCTGACAGTCAAGAGAGAGTGGCGGAGATC 990
Qy      245 -----GlyGlyGlnGluAspGlyGlySerAspGlyMetGlyGlyAspGlySerGluPhe 262
Db      991 CTCCTGCTGGAGACATGAGCGCGGCTGTGGGCTG-----ATTCCGCGC 1038
Qy      263 LeuGlnArgAspPheSerGlnThrTyrGlnArgTyrHisThrGlnSerLeuGlnAsnMet 282
Db      1039 CTGGAAAGGCGCTGCTCGACTCGCAACAGCGCGCGACAGACAGCAGCAGGAACATC 1098
Qy      283 ---SerLysGlnGluLeuLysGlyTyrLeu-----GluLeuGlu 295
Db      1099 GATGCTCTCGGCGACATGGCGCCAGAGTTGCTGGCCCTGAGACTGCCGCGGAACTGCC 1158
Qy      296 LysCysLeuSerArg----- 300
Db      1159 AAGCCGCTGAGACGATTCGCCCGCGATTCAGAGAGCGCCCGGAGTCCGCGGAGATT 1218
Qy      301 -----MetGlnAspGlnAsnAsnArgLeuArgLeuGlu-----SerLysArgLeuGly 316
Db      1219 CGGATCTGCTGTCAGCGAGCTGAGCAGGCTGCAACGCGAGCGCTGGCCGAGCGCAAGGCG 1278
Qy      317 GlyAsp-----AspAlaArgValArgGluLeu-----GluLeu 327
Db      1279 GGGAGATCGCGAAGACGATGCGCCGAGCCTGTTCACAGCTGTTCGCGGCAAGAGTGC 1338
Qy      328 GlnLeu-AspArgLeuArgAlaGlnAsnLeuGlnLeuLeuThrGlnAsnGlnLeuHisAr 347
Db      1339 GAAACCAACCGCGAGCGCGAGCGCATGCTGCTGCGGTGTGGCGGCAAGCAATACACCG 1398
Qy      347 GGGGlnGlnArgAlaPro 353
Db      1399 ATCCAAACCGCGCGCGCG 1417

```

```

RESULT 9
US-08-464-266-1
; Sequence 1, Application US/08464266
; Patent No. 5641652
; GENERAL INFORMATION:
; APPLICANT: ORO, Ph.D., ANTHONY E.
; APPLICANT: EVANS, Ph.D., RONALD M.
; TITLE OF INVENTION: INSECT RETINOID-LIKE RECEPTOR
; NUMBER OF SEQUENCES: 5
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Precty, Schroeder, Brueggemann & Clark
; STREET: 444 South Flower Street, Suite 2000
; CITY: Los Angeles
; STATE: CA
; COUNTRY: United States
; ZIP: 90071
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/464,266
; FILING DATE: 05-JUN-1995
; CLASSIFICATION: 435

```

```

; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 08/013,975
; FILING DATE: 04-FEB-1993
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 07/497,935
; FILING DATE: 22-FEB-1990
; ATTORNEY/AGENT INFORMATION:
; NAME: Reiter, Stephen E.
; REGISTRATION NUMBER: 31,192
; REFERENCE/DOCKET NUMBER: P41 9966
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 619-546-4737
; TELEFAX: 619-546-9392
; INFORMATION FOR SEQ ID NO: 1:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 2304 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: DNA (genomic)
; FEATURE:
; NAME/KEY: CDS
; LOCATION: 163..1704
;
US-08-464-266-1

```

## Alignment Scores:

```

Pred. No.: 0.00871 Length: 2304
Score: 141.00 Matches: 106
Percent Similarity: 32.43% Conservative: 49
Best Local Similarity: 22.18% Mismatches: 127
Query Match: 7.38% Indels: 197
DB: Gaps: 22

```

US-09-972-758A-2 (1-359) x US-08-464-266-1 (1-2304)

```

Qy      13 ProGlnThrSerAsnCysThrGlyAlaAlaValGlnGlnGluLeuAsnProGluArg 32
Db      388 CCCAATTCGCGCTGAGAGGCTCTCCGCGCTGAGTCCACAGACAGATCCGCTTAACCAT 447
Qy      33 ProProGlyAlaGlnGluArg-----ValProGlnGluAspSerArgTrpGlnSer 49
Db      448 CCGCTGAGCGGAGCAAGACACCTTGCTTATTTC- CGGGAGATCGGCGCAAGTGGCAAGCA 506
Qy      50 ArgAlaPheProGlnLeuGlyGlyArgProGly-----ProGluGlyGlnGly 65
Db      507 CTACGCGCTGTACAGCTGTGAGGCGCTGCAAGGCTTCTTAACGACAGTGGCGCAAGA 566
Qy      66 Ser-----LeuGlnSerGlnProProProLeuGlnThrGlnAlaCysProGlnSer 82
Db      567 TCTCACATACGCTTCAAGGAGAACCGCAATCATATACAAAGCGCGCAGAGGAA--- 623
Qy      83 SerCysLeuArgGlnGlyGlyGlyGlnLeuGlnAsnGlyAspAspSerSerLysGlyAsp 102
Db      623 ----- 623
Qy      103 PheProProAlaGluValGlnProThrProGluAlaGlnLeuLeuAlaGlnProCys 122
Db      624 -----CCGCTGCGCACTACTGCGGCTTACCAAGAGTCCCTAAGTGGCG--- 665
Qy      123 HisAspSerGlnAlaSerLysLeuGlyAla-----ProAlaAlaGlyGlyGln 138
Db      666 CATGAAGCGCGAAGCGGTCCAGAGAGAGCGTCAACGCGCGCGCCGCAATGCGGCGG--- 722
Qy      139 GlnGlnTrpGlyGlnGlnArg----- 146
Db      723 -----TAGGCTACGCGCGAGCGGCGGCGGAGCGGTCCAGGTTCCGTAAGCGGAGT 776
Qy      147 GlnLeuGlyLysLysLysHisArgArgArgProSerLysLysLysArgHisTrpLysPro 166
Db      777 CAGCTCTCAAGCGGAGAGAGAGAGAGCGG-----CGTTTCTGCGCGAGAT 821
Qy      167 TyrTyrLysLeuThrTrpGlnGlnLysLysLysPhe-AspGlnLysGlnSerLeuArgAl 186

```

```

Db 822 -----GGGAGCGGCAACGGTTCTGATGACTTCATGACCAATAGCGT 863
Qy 186 aserarg-----ileargalaglumetphealalyglu- 198
Db 864 GTCCAGGATTTTCATGACGAGGCATCATAGAGCGGAGCAGGAGCCAAATG 923
Qy 199 -----ProValAlaProTyrAsnThrThrGlnPheLeu 210
Db 924 CCGCGATCGTCACTGACGTTCTGGCGGCTTGCTCCCTATTCCACAGTCCAG----- 975
Qy 210 lAePAPhIAePAGlnGluLProAePLeuThrGlnLeu----- 224
Db 976 -----CGGACTACAAAGGTCGCTCGGCTCGGCTCGGCTCGCAAGT 1013
Qy 225 -----TyrSerLysArgAlaAlaAlaLysSerAs 234
Db 1014 GGTCAACAAACAGCTTCCAGATGTGTCAATACGGCGCATATGCCGACTTTGCCA 1073
Qy 234 pAePThrSerAspAspAspPhe----- 241
Db 1074 GGTGCGCTGAGACGACGATGATTTGCTGAAGCGGCTTGATCGAGCTGCTCATTCG 1133
Qy 242 -----MetGluGluGlyGlyGluGluAspGlyGly 252
Db 1134 GAACGTGGCTGGTGAGCATCGTTTCCTGATGACGGCGT---GCCGCGCGCGGAG 1190
Qy 252 rAePGLyMeTGLyGLyAspGLySer----- 260
Db 1191 CGGTGACTAGCCAGCATGTGCTCTTTGAGCAGCATACCGGGCTTCAGCCCAAGA 1250
Qy 261 -GluPheLuglnArgAspPheSerGlnThrGlnArgLysThrGlnSerLeu-- 279
Db 1251 GCTGTTCTCAACCAAGCTTCTCG-----TACCATCGCAACAGTGCAT 1295
Qy 279 ----- 279
Db 1296 CAAGCGGCTGTGTACCATCTTCAGACGATATTGTGAGCTGATGTAAGATGA 1355
Qy 280 ----GlnAsnMetSerLysGlnGluLeu----- 287
Db 1356 GGGGCTGAATCTTGACCGACGAGCGGCTGCTGCTGTAAGGCGCATATCTGTAACCC 1415
Qy 288 -----IleLysGlnLysLeuGluLeuGlyLysCysLeuSerArgMetGlu-- 302
Db 1416 GGACATACCGGGGATCAAGAGCGGCGGAGATCGAGATGTCGGGAGAGAGTGAACG 1475
Qy 303 -----AePGLAsnAsnArgLeuArgLeuGlnSerLysArgLeuGlyGLyAspAspAl 320
Db 1476 TTGCTGAGACGAGCACTCGCGCTGGACATCCG-----GGCGACGATCG 1520
Qy 320 sArgValAePGLyGlnGluLeuGlnLeuAspArg-LeuArgAlaGlnAsnLeuGlnLeu 340
Db 1521 AGGCTTTGGCAACTCTGCTGCTGCTGCGCGCTTTGGATGATCAAGCTGAAGTCC 1580
Qy 340 euThrGlnAsnGlnLeuLysArgGlnGlnGlnArgAlaProLeuSerLys 356
Db 1581 AGGATCACTGTTCTCTTCCGATTAACAGCAGACCGCGCTGAGAGAG 1630

```

RESULT 10  
 US-08-464-272-1  
 Sequence 1, Application US/08464272  
 Patent No. 5688691

# GENERAL INFORMATION:

APPLICANT: ORO, Ph.D., ANTHONY E.  
 APPLICANT: EVANS, Ph.D., RONALD M.  
 TITLE OF INVENTION: INSECT RETINOID-LIKE RECEPTOR  
 TITLE OF INVENTION: COMPOSITIONS AND METHODS  
 NUMBER OF SEQUENCES: 5  
 CORRESPONDENCE ADDRESS:  
 ADDRESSEE: Pretty, Schroeder, Brueggemann & Clark  
 STREET: 444 South Flower Street, Suite 2000  
 CITY: Los Angeles  
 STATE: CA

```

? COUNTRY: United States
? ZIP: 90071
? COMPUTER READABLE FORM:
? MEDIUM TYPE: Floppy disk
? COMPUTER: IBM PC compatible
? OPERATING SYSTEM: PC-DOS/MS-DOS
? SOFTWARE: PatentIn Release #1.0, Version #1.25
? CURRENT APPLICATION DATA:
? APPLICATION NUMBER: US/08/464,272
? FILING DATE:
? CLASSIFICATION: 435
? PRIOR APPLICATION DATA:
? APPLICATION NUMBER: US 08/013,975
? FILING DATE: 04-FEB-1993
? APPLICATION NUMBER: US 07/497,935
? FILING DATE: 22-FEB-1990
? ATTORNEY/AGENT INFORMATION:
? NAME: Reiter, Stephen E.
? REGISTRATION NUMBER: 31,192
? REFERENCE/DOCKET NUMBER: P41 9350
? TELECOMMUNICATION INFORMATION:
? TELEPHONE: 619-546-4737
? TELEFAX: 619-546-9392
? INFORMATION FOR SEQ ID NO: 1:
? SEQUENCE CHARACTERISTICS:
? LENGTH: 2304 base pairs
? TYPE: nucleic acid
? STRANDEDNESS: single
? TOPOLOGY: linear
? MOLECULE TYPE: DNA (genomic)
? FEATURE:
? NAME/KEY: CDS
? LOCATION: 163..1704
? US-08-464-272-1
?
? Alignment Scores:
? Pred. No.: 0.00871 Length: 2304
? Score: 141.00 Matches: 106
? Percent Similarity: 32.43% Conservative: 49
? Best Local Similarity: 22.18% Mismatches: 127
? Query Match: 7.38% Indels: 197
? DB: 1 Gaps: 22
?
? US-09-972-758A-2 (1-359) x US-08-464-272-1 (1-2304)
Qy 13 ProGlnThrSerAsnCySerThrGlyAlaAlaValGlnGluLeuAsnProGluArg 32
Db 388 CCCAATTGGCTGAGAGGCTCTGCCCGCTGCAAGCATGATCCGCTTAACCAT 447
Qy 33 ProProGlyAlaGluArg-----ValProGluGluAspSerArgTrpGlnSer 49
Db 448 CCGCTGAGCGGAGGAGCAAGACCTGCTCTATTTCGCGGATCGGGCCAGTGGCAAGA 506
Qy 50 ArgAlaPheProGlnLeuGlyLysArgProGly-----ProGluGlyGly 65
Db 507 CTACGCGGTGATACACTGCTGAGAGGCTGCAAGGCTTTCTTAACGACACATCGGCAAGA 566
Qy 66 Ser-----LeuGlnSerGlnProProLeuGlnThrGlnAlaCysProGlnSer 82
Db 567 TCTCAATACGCTTCAGGAGGAGAACCGCAATCATATAGCAAGCGGACAGAGAA--- 623
Qy 83 SerCysLeuArgGluGlyGluLysGlyGlnAsnGlyAspAspSerSerAlaGlyAsp 102
Db 623 ----- 623
Qy 103 PheProProAlaGlnValGlnLProThrProGluAlaGluLeuLeuAlaGlnProCys 122
Db 624 -----CCGCTGCGACTACTGCGCTACAGAAAGTCCTAACCTGCGG----- 665
Qy 123 HisAspSerGluAlaSerLysLeuGlyAla-----ProAlaAlaGlyGlyGlu 138
Db 666 CATGAAGCGCAAGCGGTTCAGAGAGAGGCTCAACGCGCGCGCCCAATGCGGCGG--- 722

```







```

QY 33 ProProGluAlaGluGluArg-----ValProGluGluAspSerArgTrpGlnSer 49
DB 448 CGCGTAGAGCGGAGCAAGCACTCTGCTCTATTTG--CGGGGATCGGGAGTGGGAGCA 506
QY 50 ArgAlaPheProGluGluGluArgProGly-----ProGluGluGluGly 65
DB 507 CTACGGCGGTGACAGCTGTAGAGGCTGCAGAGGGCTTTCTTAAACGACACAGTGGCGAAG 566
QY 66 Ser-----LeuGluSerGlnProProLeuGlnThrGlnAlaCysProGluSer 82
DB 567 TCTCATATACGCTTGAGGAGGAGAACCGCAACTGCATCATAGACAGCGGAGAGGAA--- 623
QY 83 SerCysLeuArgGluGluGlyGlyGlnAenGlyAspAspSerSerAlaGlyGlyAsp 102
DB 623 ----- 623
QY 103 PheProProAlaGluAlaGluProThrProGluAlaGluLeuLeuAlaGlnProCys 122
DB 624 -----CCGCTGCAGTACTGCGGCTAACAGAGTCCCTAACCTTGGCGG----- 665
QY 123 HisAspSerGlnAlaSerLysLeuGlyAla-----ProAlaAlaGlyGlyGlu 138
DB 666 CATGAAGCCGGAAGCGGCTCCAGAGAGAGCGTCAACGCGCGCGCAATGGCGGAG--- 722
QY 139 GlnGluTrpGlyGlnGlnGlnArg----- 146
DB 723 -----TAGGCTCAGCGGAGCGGAGCGGAGCGGAGTCCAGGTTCCGAGTGGAGCGGATC 776
QY 147 GlnLeuGlyLysLysLysLysLysLysLysLysLysLysLysLysLysLysLysLys 166
DB 777 CAGCTCTCAAGCGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 821
QY 167 TyrTyrLysLeuThrTrpGluGluLysLysLysPhe-AspGluLysGlnSerLeuArgAl 186
DB 822 -----GGGAGCGGAGCAAGCGTCTGATGACTTATGATACCAATAGCGCT 863
QY 186 AsnArg-----IleArgAlaGlnMetPheAlaLysGlyGln-- 198
DB 864 GTCCAGAGGATTTCTCATGAGCGGATCATAGAGCGGAGCGGAGCGGAGCGGAGCGGAG 923
QY 199 -----ProValAlaProTyrAsnThrThrGlnPheLeu 210
DB 924 CGGCGATCGTGCACTGACGTTCTCTGCGCGTGTCTCTCTATTCACAGTCCAG----- 975
QY 210 LAspAspHisAspGlnGluGluProAspLeuLysThrGlyLeu----- 224
DB 976 -----CCGAGCTAACAGAGGTCGCTGTCCGCCCTGTGCCAGT 1013
QY 225 -----TyrSerLysArgAlaAlaAlaLysSerAs 234
DB 1014 GGTCAACAAACAGCTCTTCCAGATGATGATGATGATGATGATGATGATGATGATGATGAT 1073
QY 234 PAspThrSerAspAspPhe----- 241
DB 1074 GGTGCGGTGAGACGACGAGTGTCTGCTGAAGCCGCTTGATGATGATGATGATGATGATG 1133
QY 242 -----MetGluGluGlyGlyGluGluAspGlyLys 252
DB 1134 GAACGTGCGCTGTGAGCATGCTTTCGCTGATGAGCGGGT--GCCGGCGCGCGGG 1190
QY 252 rAspGlyMetGlyGlyAspGlySer----- 260
DB 1191 CGGTGATCTAGGCGACAGATGGCTCTTTGAGGAGCATACCGGGGCTTCAGGCCAGCA 1250
QY 261 -GluPheLeuGlnArgAspPheSerLysLysLysLysLysLysLysLysLysLysLys 279
DB 1251 GGTGTTCTCTCAACCAAGAGCTTCG-----TACCATCGCAACAGTGCAGT 1295
QY 279 ----- 279
DB 1296 CAAGAGCGGTGTGTGACCATCTTGCACCGCATATTTGTGAGAGTGAAGTAAAGATGAA 1355
QY 280 ----GlnAsnMetSerLysGlnGluLeu----- 287

```

```

DB 1356 GCGGTGATCTCGACCGAGCGAGCTCTCTGTTGAAGCCATCATCTGTACACCC 1415
QY 288 -----IleLysGluTyrLeuGluGluLysCysLeuSerArgMetGlu-- 302
DB 1416 GGACATACCGGGAGTCAAGAGCGCGGAGGATGATGATGATGATGATGATGATGATGATG 1475
QY 303 -----AspGluAsnAsnArgLeuArgLeuGluSerLysArgLeuGlyGlyAspAspAl 320
DB 1476 TTGCTGAGCGAGCATCGCGCTGGAACATCGG-----GGCGAGCATGG 1520
QY 320 ArgValArgGluGluGluGluGluGluGluGluGluGluGluGluGluGluGluGluGlu 340
DB 1521 ACGCTTTCGCAACACTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 1580
QY 340 eutThrGluAsnGluLeuHisArgGlnGlnGluArgAlaProLeuSerLys 356
DB 1581 AGATACCTGCTCTCTTCCGATTCACGAGCGCGCGCTGAGAGAG 1630

```

## RESULT 13

US-09-265-013-2  
Sequence 2, Application us/09265013

Patent No. 6451304  
GENERAL INFORMATION:

APPLICANT: FRIEDMAN, Theodore  
TITLE OF INVENTION: METHOD FOR RETROVIRUS VECTOR PRODUCTION BY SEPARATED  
TITLE OF INVENTION: GAG AND POL EXPRESSION  
FILE REFERENCE: 041673/2010  
CURRENT APPLICATION NUMBER: US/09/265,013  
CURRENT FILING DATE: 1999-03-09  
NUMBER OF SEQ ID NOS: 5  
SOFTWARE: PatentIn Ver. 2.1  
SEQ ID NO 2  
LENGTH: 1617  
TYPE: DNA  
ORGANISM: Moloney murine leukemia virus (MOLMV)

US-09-265-013-2

## Alignment Scores:

Pred. No.: 0,0117 Length: 1617  
Score: 137.00 Matches: 90  
Percent Similarity: 30.63% Conservative: 42  
Best Local Similarity: 20.88% Mismatches: 152  
Query Match: 7,17% Indels: 147  
DB: 4 Gaps: 17

US-09-972-758A-2 (1-359) x US-09-265-013-2 (1-1617)

```

QY 18 CyThrGly-----AlaAlaAlaValGlnGluLeuAsnProGluArgProPro-GlyAl 36
DB 300 TGTACACCTTAAGCCTTCGCGCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCT 359
QY 36 aGlnGluArgValProGluGluAspSerArgTrpGlnSerAlaGalaPheProGlnLeu 56
DB 360 TCCCTGTTGACCGCGCTCGATCTCTCTTATCCAGCCCTCACT---CTTCTCTAGG 416
QY 56 yGlyArgProGlyProGlu-----GlyGluGlySer-----LeuG 68
DB 417 CGCCAAACCTTAACCTCAAGTTCTTTCGACAGTGGGGGCGCTCATCGACCTACTTAC 476
QY 68 userGlnProProLeuGlnThrGlnAlaCysProGluSerSerCysLeuArgGlu 88
DB 477 AGAAGACCCCGCTTATAGGAGCCCAAGACACCTCTTCCAGC-----AGGAGCG 530
QY 88 yGluLysGlnAenGlyAspAspSerSerAlaGlyGlyAspPheProProProAlaG 108
DB 531 A-----AATGATGAGGAAGAGCACTTCGAGGAGGAGGAGGAGGAGGAGGAGGAG 567
QY 108 uValGluProThrProGluAlaGluLeu-----AlaGlnProCysHis 124
DB 568 ----GACCCCTCCCAATGGATCTTCGCTTACGTGGAGAGCGGAGCGCCCTGTGGCGGA 623

```

```

QY 124 pSerGluAlaSerLeuLeuGluAlaProAlaAlaGlyGluGluGluTrpGluGlu 144
DB 624 CTCCTACTCTCCGAGCATTTCCCTCCGCGAGAGGAAAC----- 666
QY 144 ngInArgGlnLeuGlyLysLysLysHisArgArgArgProSerLysLysArgHisTr 164
DB 667 -----GGACAGCTTCATATCTGCGCGCTTCTCTCTTGCAGCTTTACACTG 713
QY 164 pLysProTyTyTyLysLeuThrTrpGluGluLysLysLysPheAspGluLysGlnSerLe 184
DB 714 GAAA-----AATATAACCTCTTTCTTTCGAAATCCAGGTAA 752
QY 184 uArgAlaSerArgLLeuArgAlaGluMetPheAlaLysGlyGlnProValAlaProTyTrs 204
DB 753 ACTGACAGCTCTGATCGAGTCTGTCTCTATC----- 783
QY 204 nThrTrgInPheLeuMetAspAspHisAspGlnGluGluProAspLeuLysThrGly-- 223
DB 784 -ACCCATCGACCCCACTGGAGCAGCTGTACGACGCTTGGGAGACTGTGTCACCGAGA 842
QY 224 -----LeuTySerLysArgAlaAlaAlaLysSerAspAsp----- 235
DB 843 AGAAAAACAACGGGTGCTCTTAAGCTAGAAAGCGGTGCGGCGCATGTGGCGGCC 902
QY 236 -----ThSerAspAspAsp 241
DB 903 CACTCAATGCGCAATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 962
QY 241 eMetGluGluGlyGlyGluGlu----- 248
DB 963 CACCAACCGAGGAGTAGAACCACTATGCACTATGCGCACTGCTCTTACCGGCTCT 1022
QY 249 -----AspGlyGlySerAspGly 255
DB 1023 CCAAAAACGGGAGAGAAACCAATTTGGCCAGTAAAGAAATACAAAGAGGCC 1082
QY 255 rGlyGlyAspGlySerGluPheLeuGlnArgAspPheSerGluTrgGluArgTyTr-- 274
DB 1083 CATGAGTCTCTCGCGCTTCTCTAGAGAGAA---CTTAAGGAAGCCATGCGAGTACAC 1139
QY 274 ----- 274
DB 1140 TCCTATGACCTGAGAGACCCAGGAGCAAACTAATGTCTATGCTTTCTTACGCA 1199
QY 275 -----HisTrgGluSerLeuGlnAspMetSerLys 284
DB 1200 GTCTGCCCGAGCATTTGGGAGAAAGTTAGAGAGTTAGAAATTTAAATAACAAGCGCT 1259
QY 284 rGlnGluLeuLLeuGlyLysGluTrgLeuGluGluLysCysLeuSerArgMetGluAspG 304
DB 1260 TGGAGATTGGTTAGAGAGAGCAAAAGATCTTAATAAAGCAAAACCCCGAGAGAAAG 1319
QY 304 uAsnAsnArgLeuArgLeuGlnSerLysArgLeuGlyGlyAspAspAlaArgValArgG 324
DB 1320 AGAGGAACGTATCAGG-----AGAG 1340
QY 324 uLeuGlnLeuGlnLeuAspArgLeuArgAlaGluAsnLeuGlnLeuThrGlnAsnG 344
DB 1341 AACAAGGAAAAAAGAAAGACCGCTAGACAGAGATGACAAAGAAAGAAAGAGAA 1400
QY 344 uLeuHisArgGlnGlnGluArgAlaProLeu 354
DB 1401 TCCTAGAGAGCATAGAGATGAGCAAGCTA 1431

```

```

; TITLE OF INVENTION: Expression systems
; FILE REFERENCE: 09/011,745
; CURRENT APPLICATION NUMBER: US/09/011,745
; EARLIER FILING DATE: 1998-06-22
; EARLIER APPLICATION NUMBER: PCT/GB96/02061
; EARLIER FILING DATE: 1996-08-23
; EARLIER APPLICATION NUMBER: GB9517263.1
; EARLIER FILING DATE: 1995-08-23
; NUMBER OF SEQ ID NOS: 29
; SOFTWARE: Patent Ver. 2.0
; SEQ ID NO 3
; LENGTH: 7308
; TYPE: DNA
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: Description of Artificial Sequence: Portion of
; OTHER INFORMATION: construct
US-09-011-745-3

Alignment Scores:
Pred. No.: 0.0938 Length: 7308
Score: 137.00 Matches: 90
Percent Similarity: 30.63% Conservative: 42
Best Local Similarity: 20.88% Mismatches: 152
Query Match: 7.17% Indels: 147
DB: 3 Gaps: 17

US-09-972-758A-2 (1-359) x US-09-011-745-3 (1-7308)

QY 18 CyThrGly---AlaAlaAlaValGlnGluLeuAsnProGluArgProPro-GlyVal 36
DB 1866 TGTACACCTTAAGCCCTCCGCTCTTCTCTCCTACCCGCGCTCTTCCCTGGAACC 1925
QY 36 aGluGluArgValProGluGluAspSerArgTrpGlnSerArgAlaPheProGlnLeuG 56
DB 1926 TCCTGTTGACCCCGCTCGATCTCTTATTCAGCCCTCACT---CTTCTCTAGG 1982
QY 56 yGlyArgProGlyProGlu-----GlyGluGlySer-----LeuG 68
DB 1983 CGCAAACTTAACCTCAAGTTCTTTCTGACAGTGGGGGGCGCTCATCGACCTTATC 2042
QY 68 uSerGlnProProProLeuGlnThrGlnAlaCysProGluSerSerCysLeuArgGluG 88
DB 2043 AGAAGACCCCGCTTATAGGAGCCCAAGACACCCCTTCCGAC---AGGACCG 2096
QY 88 yGluLysGlyGlnAsnGlyAspAspSerSerAlaGlyGlyAspPheProProAlaG 108
DB 2097 A-----AATGTGGAGAAAGCAACCTTGGGAGAGCAGCG----- 2133
QY 108 uValGluProThrProGluAlaGlnLeuLeu-----AlaGlnProCysHisAs 124
DB 2134 ---GACCCCTCCCAATGAGCATCTCGCTACGTGGAGACGGAGCCCTGTGGCGCA 2189
QY 124 pSerGluAlaSerLysLeuGlyAlaProAlaAlaGlyGlyGluGluGluTrpGlyGln 144
DB 2190 CTCCTACTCTCCGAGCATTTCCCTCCGCGAGAGGAAAC----- 2232
QY 144 ngInArgGlnLeuGlyLysLysLysHisArgArgArgProSerLysLysArgHisTr 164
DB 2233 -----GGACAGCTTCATATCTGCGCGCTTCTCTCTTGCAGCTTTACACTG 2279
QY 164 pLysProTyTyTyLysLeuThrTrpGluGluLysLysLysPheAspGluLysGlnSerLe 184
DB 2280 GAAA-----AATATAACCTCTTTCTTTCGAAATCCAGGTAA 2318
QY 184 uArgAlaSerArgLLeuArgAlaGluMetPheAlaLysGlyGlnProValAlaProTyTrs 204
DB 2319 ACTGACAGCTCTGATCGAGTCTGTCTCTATC----- 2349
QY 204 nThrTrgInPheLeuMetAspAspHisAspGlnGluGluProAspLeuLysThrGly-- 223
DB 2350 -ACCATCGACCCCACTGGAGCAGCTGTGACGAGCTGTTGGGAGACTGTGTCAGCGAGA 2408

```

```

QY 224 -----LeuTyrSerIysArgAlaAlaIAluSerAspAsp----- 235
Db 2409 AGAAAAACAACGCGTCTCTTAGAGCGTGAAGGCGGCGGATGATGCGGCC 2468
QY 236 -----ThSerAspAspAspPh 241
Db 2469 CACTCACTGCCCAATGAAGTGATGCGGCTTTTCCCTCGAGCGCCAGACTGGGATTA 2528
QY 241 eMetGluGluGluGluGlu----- 248
Db 2529 CACCAACCCAGGAGAGAGAACCACTAGTCCACTATGCCAGTTGCTCTAGCGGCTCT 2588
QY 249 -----AspGlyGlySerAspGlyWe 255
Db 2589 CCAAAACGGGCGAGAGCCCAACCAATTTGGCCAGAGTAAAGAAATACACAAAGGCC 2648
QY 255 tGlyGlyAspGlySerGluPheLeuGluInArgAspPheSerGluThrTyrGluArgTyr-- 274
Db 2649 CAATGAGTCTCCCTCGGCTCTCTAGAGAGA---CTTAAGAAAGCCTATCGAGGTACAC 2705
QY 274 ----- 274
Db 2706 TCCTTATGACCTGAGAGACCCAGGCAAGAACTAATGTCTATGCTTTTCATTGGCA 2765
QY 275 -----HisThrGluSerLeuGluAsnMetSerIy 284
Db 2766 GTCTGCCCGCAGACATTGGAGAAAGTTAGAGAGTTAAAGATTAAACAAAGCGCT 2825
QY 284 sGluGluLeuIleIysGluTyrLeuGluLeuGluIlyScySLeuSerArgMetGluAspG1 304
Db 2826 TGGAGATTGTTGTTAGAGAGCAGAAAGATCTTAAATMAACGAGAAACCCCGAAGAAAG 2885
QY 304 uAsnAspArgLeuArgLeuGluSerIysArgLeuGluGlyAspAspAlaArgValArgG1 324
Db 2886 AAGAGAACCTATCAG-----AGAGA 2906
QY 324 uLeuGluLeuGluLeuAspArgLeuArgAlaGluAsnLeuGluLeuThrGluAsnG1 344
Db 2907 AACAGAGAAAGAAAGAAAGCGCGTAGACAGAGATGACAGAAAGAAAGAAAGAGA 2966
QY 344 uLeuHisArgGluGluGluArgAlaProLeu 354
Db 2967 TCGTAGAGACATAGAGATGAGCAAGCTA 2997

RESULT 15
US-09-011-745-4
/ Sequence 4, Application US/09011745
/ Patent No. 6165715
/ GENERAL INFORMATION:
/ APPLICANT: Collins, Mary KL
/ APPLICANT: Weiss, Robin A
/ APPLICANT: Takeuchi, Yasuhiro
/ APPLICANT: Cosset, Francois-Loic
/ TITLE OF INVENTION: Expression systems
/ FILE REFERENCE: 09/011,745
/ CURRENT APPLICATION NUMBER: US/09/011,745
/ EARLIER FILING DATE: 1998-06-22
/ EARLIER APPLICATION NUMBER: PCT/GB96/02061
/ EARLIER FILING DATE: 1996-08-23
/ EARLIER APPLICATION NUMBER: GB9517263.1
/ NUMBER OF SEQ ID NOS: 29
/ SOFTWARE: PatentIn Ver. 2.0
/ SEQ ID NO 4
/ LENGTH: 7308
/ TYPE: DNA
/ ORGANISM: Artificial Sequence
/ FEATURE:
/ OTHER INFORMATION: Description of Artificial Sequence: Portion of
/ US-09-011-745-4
Alignment Scores:

```

```

Pred. No.: 0.0938 Length: 7308
Score: 137.00 Matches: 90
Percent Similarity: 30.63% Conservative: 42
Best Local Similarity: 20.88% Mismatches: 152
Query Match: 7.17% Indels: 147
DB: 3 Gaps: 17

US-09-972-758a-2 (1-359) x US-09-011-745-4 (1-7308)
QY 18 CysThrGly--AlaAlaAlaValGluGluLeuAsnProGluIlyArgProPro-GlyAl 36
Db 1866 TGTACACCTTAAGCCTTCGCGCTCTCTCTCTCATCCGCCCTCTCTCTCTCTCTTGAACC 1925
QY 36 aGluGluArgValProGluGluAspSerArgTyrIleAsnArgAlaPheProGluLeuG1 56
Db 1926 TCCTCGTTGACCCCGCGCTCGATCTCTCTTATCAGCCCTCAGT---CTTCTCTAGG 1982
QY 56 yGlyArgProGlyProGlu-----GlyGluGlySer-----LeuG1 68
Db 1983 CGCCAAACCTMAACCTCAAGTTCTTTCGACAGTGGGGGCGCTCATGACCTACTTAC 2042
QY 68 userGluProProProLeuGluInThrGluAlaCysProGluSerSerCysLeuArgGluG1 88
Db 2043 AGAAGACCCCGCGCTTATAGGACCAAGACACCCCTTCCAGC-----AGGAGCG 2096
QY 88 yGluIySglIyGluAsnGlyAspAspSerSerIaGlyGlyAspPheProProAlaG1 108
Db 2097 A-----AATGTGAGAAAGAGCAGCCCTCGGAGAGGACCG----- 2133
QY 108 uValGluProThrProGluAlaGluLeuLeu-----AlaGluProCysHisAs 124
Db 2134 ---GACCCCTCCCAATGCGATCTCGCTACGTGGAGACCGGAGCCCTCTGTGGCGCA 2189
QY 124 pSerGluIleSerIysLeuGlyAlaProAlaIleGlyGlyGluGluGluTyrGlyGluG1 144
Db 2190 CTCACACTCTCGAGGCAATTCCTCCCGCCAGAGAGAAAC----- 2232
QY 144 nGluArgGluLeuGlyIlySlyIysHisArgArgArgProSerIySlyIysArgHisI 164
Db 2233 -----GGAACGCTTCAATACCTGCGCGCTTCTCTCTTTCGACCTTACACAG 2279
QY 164 pIysProTyrTyrIlySlyLeuThrTyrGluGluIlySlyIysPheAspGluIyGluSerI 184
Db 2280 GAAA-----AATTAATACCTTCTTTTTCGAAATCCAGATCAAGTAA 2318
QY 184 uArgAlaSerArgIleArgAlaGluMetPheAlaIySglIyGluInProValAlaProTyrAs 204
Db 2319 ACTGACAGCTCTGATCGAGTCTGTTCTCATC----- 2349
QY 204 nThrThrGluPheLeuMetAspAspHisAspGluGluGluProAspLeuIyThrGly-- 223
Db 2350 -ACCATCAAGCCCACTGGAGAGCATGTCAGAGCTGTGGGAGACTCTCTGACCGGAGA 2408
QY 224 -----LeuTyrSerIysArgAlaAlaIAluSerAspAsp----- 235
Db 2409 AGAAAAACAACGCGTCTCTTAGAGCGTGAAGGCGGCGGATGATGCGGCC 2468
QY 236 -----ThSerAspAspAspPh 241
Db 2469 CACTCACTGCCCAATGAAGTGATGCGGCTTTTCCCTCGAGCGCCAGACTGGGATTA 2528
QY 241 eMetGluGluGluGluGlu----- 248
Db 2529 CACCAACCCAGGAGAGAGAACCACTAGTCCACTATGCCAGTTGCTCTAGCGGCTCT 2588
QY 249 -----AspGlyGlySerAspGlyWe 255
Db 2589 CCAAAACGGGCGAGAGCCCAACCAATTTGGCCAGAGTAAAGAAATACACAAAGGCC 2648
QY 255 tGlyGlyAspGlySerGluPheLeuGluInArgAspPheSerGluThrTyrGluArgTyr-- 274
Db 2649 CAATGAGTCTCCCTCGGCTCTCTAGAGAGA---CTTAAGAAAGCCTATCGAGGTACAC 2705

```

```
Qy 274 ----- 274
Db 2706 TCCTTATGACCTGAGAGCCAGGCAAGAACTAATGTCATATCTTTCATTGGCA 2765
Qy 275 -----HistHrGluSerLeuGluAsnMetSerLy 284
Db 2766 GTCTGCCCAAGACATTGGGAGAAAGTTAGAGAGCTTAGAAGATTAAACAAACAGACGCT 2825
Qy 284 sGInGluLeuIleLyGluTyrLeuGluLeuGluLysCysLeuSerArgMetGluAspG1 304
Db 2826 TGGAGATTGGTTAGAGAGCGCAGAAAGATCTTAATAACGAGAAACCCGGAGAGAAAG 2885
Qy 304 uAsnAsnArgLeuArgLeuGluSerLysArgLeuGlyGlyAspAspAlaArgValArgG1 324
Db 2886 AGAGGAACGTCATCAGG-----AGAGG 2906
Qy 324 uLeuGluLeuGluLeuAspArgLeuArgAlaGluAsnLeuGluLeuLeuThrGluAsnG1 344
Db 2907 AACAGAGAGAAAGAAAGAACGCCGTAGGACAGAGATGAGCAGAAAGAGAAAGAGAGA 2966
Qy 344 uLeuHisArgGInGInGluArgAlaProLeu 354
Db 2967 TCCTAGAGAGCATAGAGAGATGAGCAAGCTA 2997
```

Search completed: February 5, 2004, 15:52:30  
Job time : 124 secs

GenCore version 5.1.6  
Copyright (c) 1993 - 2004 Compugen Ltd.

OM protein - nucleic search, using frame\_plus\_p2n model

Run on: February 5, 2004, 13:32:05 ; Search time 395 Seconds  
(without alignments)  
2451.413 Million cell updates/sec

Title: US-09-972-758A-2  
Sequence: 1 MAEPFLSEYQHPQTSNCTG.....LTENELHQERAPLSKFGD 359

Scoring table: BLOSUM62  
Xgapop 10.0 , Xgapext 0.5  
Ygapop 10.0 , Ygapext 0.5  
Fgapop 6.0 , Fgapext 7.0  
Delop 6.0 , Delext 7.0

Searched: 2552756 seqs, 1349719017 residues  
Total number of hits satisfying chosen parameters: 5105512

Minimum DB seq length: 0  
Maximum DB seq length: 200000000  
Post-Processing: Minimum Match 0%  
Maximum Match 100%  
Listing first 45 summaries

Command line parameters:  
-MODE=frame\_plus\_p2n.model -DEV=xlh  
-Q=/cgn2\_1/USPTO.epool/US09972758/runat\_05022004\_095002\_1181/app\_query.fasta.1.519  
-DB=N\_Geneseq\_19Jun03 -QFMT=fastap -SUFFIX=ring -MINMATCH=0.1 -LOOPEL=0  
-LOOPEXT=0 -UNITS=bits -START=1 -END=1 -MATRIX=biosum62 -TRANS=human40.cdi  
-LIST=45 -DOCALLIGN=200 -THR SCORE=pct -THR MAX=100 -THR MIN=0 -ALIGN=15  
-MODE=LOCAL -OUTFMT=ptc -NORM=ext -HEAPSIZ=500 -MINLEN=0 -MAXLEN=2000000000  
-USER=US09972758.QCEN\_1\_1.0@runat\_05022004\_095002\_1181 -NCPU=6 -ICPU=3  
-NO\_MMAP -LARGEQUERY -NEG SCORES=0 -WAIT -DSPELCK=100 -KONJLOG  
-DEV TIMEOUT=120 -WARN TIMEOUT=30 -THREADS=1 -XGAPOP=10 -XGAPEXT=0.5 -FGAPOP=6  
-FGAPEXT=7 -YGAPOP=10 -YGAPEXT=0.5 -DELOP=6 -DELEXT=7

Database :

N\_Geneseq\_19Jun03:\*  
1: /SIDSI/gcgdata/geneseq/geneseqn-emb1/NA1980.DAT:\*  
2: /SIDSI/gcgdata/geneseq/geneseqn-emb1/NA1981.DAT:\*  
3: /SIDSI/gcgdata/geneseq/geneseqn-emb1/NA1982.DAT:\*  
4: /SIDSI/gcgdata/geneseq/geneseqn-emb1/NA1983.DAT:\*  
5: /SIDSI/gcgdata/geneseq/geneseqn-emb1/NA1984.DAT:\*  
6: /SIDSI/gcgdata/geneseq/geneseqn-emb1/NA1985.DAT:\*  
7: /SIDSI/gcgdata/geneseq/geneseqn-emb1/NA1986.DAT:\*  
8: /SIDSI/gcgdata/geneseq/geneseqn-emb1/NA1987.DAT:\*  
9: /SIDSI/gcgdata/geneseq/geneseqn-emb1/NA1988.DAT:\*  
10: /SIDSI/gcgdata/geneseq/geneseqn-emb1/NA1989.DAT:\*  
11: /SIDSI/gcgdata/geneseq/geneseqn-emb1/NA1990.DAT:\*  
12: /SIDSI/gcgdata/geneseq/geneseqn-emb1/NA1991.DAT:\*  
13: /SIDSI/gcgdata/geneseq/geneseqn-emb1/NA1992.DAT:\*  
14: /SIDSI/gcgdata/geneseq/geneseqn-emb1/NA1993.DAT:\*  
15: /SIDSI/gcgdata/geneseq/geneseqn-emb1/NA1994.DAT:\*  
16: /SIDSI/gcgdata/geneseq/geneseqn-emb1/NA1995.DAT:\*  
17: /SIDSI/gcgdata/geneseq/geneseqn-emb1/NA1996.DAT:\*  
18: /SIDSI/gcgdata/geneseq/geneseqn-emb1/NA1997.DAT:\*  
19: /SIDSI/gcgdata/geneseq/geneseqn-emb1/NA1998.DAT:\*  
20: /SIDSI/gcgdata/geneseq/geneseqn-emb1/NA1999.DAT:\*  
21: /SIDSI/gcgdata/geneseq/geneseqn-emb1/NA2000.DAT:\*  
22: /SIDSI/gcgdata/geneseq/geneseqn-emb1/NA2001A.DAT:\*  
23: /SIDSI/gcgdata/geneseq/geneseqn-emb1/NA2001B.DAT:\*  
24: /SIDSI/gcgdata/geneseq/geneseqn-emb1/NA2002.DAT:\*  
25: /SIDSI/gcgdata/geneseq/geneseqn-emb1/NA2003.DAT:\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed.

and is derived by analysis of the total score distribution.

# SUMMARIES

Result No.	Score	Query Match	Length	DB	ID	Description
1	1910	100.0	2199	19	AAV82778	Clone bp783_3 isol
2	1910	100.0	2199	24	ABQ92015	Human polynucleoti
3	1903	99.6	2178	22	AAH18098	Human cDNA sequenc
4	1902	99.6	1080	24	ABN84013	Human osteogen do
5	733	38.4	1083	24	ABL90067	Human polynucleoti
6	714	37.4	523	20	AAV86036	EST clone B227. H
7	691	36.2	414	25	ABX54966	Bovine EST associa
8	629	32.9	461	24	ABL63543	Breast cancer rela
9	545.5	28.6	1557	24	ABQ61099	Mouse putative pro
10	308.5	16.2	997	22	AAH99667	Human protein enco
11	245	12.8	342	20	AAV87316	EST clone BP783.
12	214	11.2	1489	23	ABL06282	Drosophila melanog
13	205.5	10.8	3714	23	AAV95217	Drosophila melanog
14	184	9.6	549	21	AAV95217	Cat flea head and
15	174	9.1	1024	23	AAV79493	DNA encoding novel
16	162	8.5	7296	24	ABT08505	Human novel protei
17	162	8.5	7516	22	AAK51987	Human polynucleoti
18	159.5	8.4	4041	24	ABQ54663	Human ovarian anti
19	159.5	8.4	7726	23	AAK55910	DNA encoding novel
20	158.5	8.3	9651	20	AAZ22301	cDNA encoding a hu
21	156	8.2	6432	23	AAK55903	DNA encoding novel
22	155	8.1	2277	19	AAV13836	Homo sapiens mamma
23	155	8.1	2277	19	AAV05372	Human telomerase p
24	153.5	8.0	3579	23	ABL05651	Drosophila melanog
25	153.5	8.0	6105	23	ABL05650	Drosophila melanog
26	152	8.0	2421	24	ABK80600	Bacillus clausii g
27	152	8.0	3540	23	ABL06305	Drosophila melanog
28	151	7.9	3707	23	ABL06276	Drosophila melanog
29	149.5	7.8	8486	22	AAK52971	Human polynucleoti
30	149	7.8	2483	23	ABL07559	Drosophila melanog
31	149	7.8	4483	23	ABL07558	Drosophila melanog
32	149	7.8	18737	23	ABL07492	Drosophila melanog
33	149	7.8	18737	23	ABL07530	Drosophila melanog
34	148.5	7.8	7741	22	AAI60155	Human polynucleoti
35	148.5	7.8	7741	22	AAI60156	Human polynucleoti
36	148	7.7	5943	22	AAK22920	Lung cancer associ
37	147.5	7.7	3773	21	AAI18028	Drosophila melanog
38	147.5	7.7	15606	23	ABL16881	Human polynucleoti
39	147	7.7	7453	22	AAI58369	Human polynucleoti
40	147	7.7	7501	22	AAI58370	Human polynucleoti
41	145	7.6	4301	24	ABK63756	Rat sequence diffe
42	145	7.6	8731	23	ABL10296	Drosophila melanog
43	144	7.5	4714	22	ABD21393	Human Cockayne syn
44	143.5	7.5	2159	15	AAO67223	Mouse p55Nuc. Mus
45	143	7.5	867	23	AAK84549	DNA encoding novel

## ALIGNMENTS

RESULT 1  
AAV82778  
ID AAV82778 standard; cDNA; 2199 BP.  
XX  
AC AAV82778;  
XX  
DT 25-FEB-1999 (first entry)  
XX  
DE Clone bp783\_3 isolated from human foetal kidney cDNA library.  
XX  
XX secreted protein; nutritional activity; immune stimulating; vaccine;  
XX suppressing activity; haematopoiesis regulating activity;  
XX tissue growth activity; activin; inhibin activity; chemotactaxis;  
XX chemokinetic activity; haemostasis; thrombolytic activity; receptor;  
XX ligand; anti-inflammatory; cadherin; tumour invasion suppressor;  
XX tumour inhibition; gene therapy; ds.  
XX

XX	Homo sapiens.
XX	MO98642739-A2.
PN	01-OCT-1998.
PD	
XX	
XX	20-MAR-1998; 98WO-US05653.
XX	
PR	19-MAR-1998; 98US-0044466.
ER	21-MAR-1997; 97US-0822167.
PA	(GENY ) GENETICS INST INC.
PI	Agostino MJ, Jacobs K, Lavallie ER, McCoy JM, Merberg D;
PI	Racie LA, Spaulding V, Treacy M;
DR	WP1; 1998-609890/51.
DR	P-PSDB; AAW85455.
PT	New polynucleotides encoding secreted human proteins - derived from
PT	human foetal brain, adult brain, foecal kidney, placenta or adult
PT	pineal gland cDNA libraries.
XX	
PS	Claim 1; Page 66-67; 113pp; English.
CC	
CC	The present sequence encodes a secreted protein. The polynucleotide and
CC	secreted protein are predicted to have biological activities which would
CC	make them suitable for treating, preventing or ameliorating medical
CC	conditions in humans and animals, although no supporting data is given.
CC	Secreted activities include nutritional activity, immune stimulating
CC	(e.g. as vaccines) or suppressing activity, haematopoiesis regulating
CC	activity, tissue growth activity, activin/inhibin activity,
CC	chemotherapeutic/chemokinetic activity, haemostatic and thrombolytic activity,
CC	receptor/ligand activity, anti-inflammatory activity, cadherin/tumour
CC	invasion suppressor activity, and tumour inhibition activity (no data is
CC	given in the specification to support these activities). The
CC	polynucleotide is also stated to be useful for gene therapy.
XX	
SO	Sequence 2199 BP; 552 A; 511 C; 674 G; 462 T; 0 other;
Alignment Scores:	
Pred. No.:	8,186-106 Length: 2199
Score:	1910.00 Matches: 359
Percent Similarity:	100.00% Conservative: 0
Best Local Similarity:	100.00% Mismatches: 0
Query Match:	100.00% Indels: 0
Gaps:	19 Gaps: 0
US-09-972-758A-2 (1-359) x AAV82778 (1-2159)	
QY	1 MetalagIUPropheLeuSerGluTrgInhiSginProGlnThrSerAsnCyThrGly 20
Dd	707 ATGGCGGACCATTTCTGTGCAGAAATATCAACACCAGCCTCAAATACGTACAGT 766
YY	21 ALAAlaAlaValGlnGlnGlnLuleuAsnProGIuaGrProProGlyAlAGlnGlnLyrAl 40
Dd	767 GCTGTGTGTGTCACAGAAAGAGTGTAACCTGAGCCGCCCCACGGGCGAGAGGGGGT 826
QY	41 ProGlnGluAspSerArgTrGlnSerArgAlAPheProGlnLueGlyGlyArpProGly 60
Dd	827 CCCGAGGAGACAGTAGTGTGGCAATCGAGAAGCGTTCCCCACTTGGGTGGCCGTCCGG 886
QY	61 ProGlnGlyGlnGlySerLeuGlnuSerGlnProProProLeuGlnThrGlnAlaCyAsPro 80
Dd	887 CCGAGGGGGGAAGGAGAGCTGGAAATCCCAACCACTCCCTTSCAACCCAGGCGTGTCCA 946
QY	81 GluSerSerCyGleuArGlnGlnGlyGluLysGlyGlnAsnGlyAspAspSerSerAlaGly 100
Dd	947 GAATTACTGCTGTCAGAGAGGCGCAGAAAGGGCCGAATGGGGAGACATCTCTCCCTGGC 1006
QY	101 GLyApPHeProProProAlaGlnLyrAlGlnProThrProGlnLyrAlAGlnLueLueLagln 120
Dd	1007 GGCGACTTCCGCCGCCCGCACAGATGGAAACCGACGCCCGAGGCCGAGCTGCTGCCCGAG 1066

Oy		12	Procyoniaspergulariaserlyslengualaprolalaglyglugluglu	140
Db		1067	CCTTGTATTAATCCCGAGGCCAATAAGTTGGGGGCTCCTCCGACGGGGCGGAAGAAG	1126
Oy		141	TTPGLYGLNGLNARGLNLEUGLYLVALYSIASHARGARGLPROSERLYLS	160
Db		1127	TGGGACAGACAGACAGACTGGGGAAAAAAAATCATGAGAGACCCTGCCTCAAGAG	1186
Oy		161	Lysarghistridylserprotyrtyrtylvalyleuthtrigluilulyslylslysphenaspglu	180
Db		1187	AAGGGGATTGGAAACCGTACTCACACTGACCTGGGAAAGAAAAGTTTCAGCAG	1246
Oy		181	Lysginserleuaragalaserarglylearylaglunetphealalysglyglinproval	200
Db		1247	AAACAAGACCTTGAGCTTCAGAGATCCGAGCGAGATGTTCCGCAAGGGCCAGCGGTC	1306
Oy		201	Alaprotyranthrthrghlnphelemecaspasphisapnglngluiproaspheu	220
Db		1307	GCGCCCTATTAACACACAGCAGTTCCTCATGATGATACACACAGGAGGAGCCGGAATCC	1366
Oy		221	Lysethrclyleuyrserlysaargalaialalyseraspasprthrsesarpaspasp	240
Db		1367	AAACCCGGCTGTACTCCAAGCGGGCCGCCCAAAATCCACACGACACCGATGACGAC	1426
Oy		241	PheMETGLUGLUGLYGLYGILUASPGLYGLYSEZASPLGMETGLYGLYASPGLYSER	260
Db		1427	TTCAATGSAABAAGGGGGGTGAAGAGATGGGGGACGCATGGATGGAGGGGACCGCAGC	1486
Oy		261	GlupheulglnnaraspPheserglutThrTyrglunargtryrhsthrglusertleugn	280
Db		1487	GAGTTTCTGACGGCGGACTTCTCGGAGACCTACGAGCGTACACACGAGAGCGTCTCAG	1546
Oy		281	AsmserterysglnnglualeuilelysgluTYrtleugluenguilyusCyalseserArg	300
Db		1547	AACATGACCAAGCAGGAGGCTCATACAGAGATCCTGGAACTGGAGAACTGCTCTCCGC	1606
Oy		301	Mecgluaspglunasbnanargleuargleugluserylsargtleuglygl yasppalpa	320
Db		1607	ATGAGAGACGACGAACAACCGGCTGCGGCTGGAGAGCAAGCGGCTGGGTGGACAGACGCG	1666
Oy		321	ArgValaIrsgluleugluLeugluLeuaspargleuarglaaglunleuglnleulleu	340
Db		1667	CGTGTGCGGAGGCTGGAGCTGGAGCTGGACCGGCTGCGCCCGACGAACCTCCAGCTGCTG	1726
Oy		341	ThrglunasngluenuhsargnglnglnlarArAlaproleserlyspPhegiYasp	359
Db		1727	ACCGAGAACAACTGCACCGGACGACGACGAGCGCCGCTTTCANAATTGGAGAC	1783
<b>RESULT 2</b>				
ID	ABQ92015		standard; cDNA; 2199 BP.	
XX	ABQ92015;			
DE	04-OCT-2002	(first entry)		
DT	Human polynucleotide SEQ ID NO 12.			
XX				
XX				
KM	Human, cytostatic; antirheumatic, antiarthritic; vulnerary; analgesic,			
KM	antiinflammatory; antibacterial; immunosuppressive; antiparkinsonian;			
KM	neuroprotective; nootropic; osteopathic; haemostatic; vasodropic;			
KM	antitumor; fungicide; antidiabetic; antischismatic; anticlergalic;			
KM	immunostimulant; antiparasitic; secreted protein; transmembrane protein;			
KM	cytokine; cell proliferation; cell differentiation; autoimmune disease;			
KM	stem cell; growth factor; nervous system disease; neuropathy;			
KM	Alzheimer's disease; Parkinson's disease; Huntington's disease;			
KM	osteoporosis; severe combined immunodeficiency; SCID; infection;			
KX	multiple sclerosis; rheumatoid arthritis; gene therapy; gene; ss.			
OS	Homo sapiens.			
XX				
NN	US2002065394-A1.			

XX 30-MAY-2002.  
PD 22-DEC-2000; 2000US-0745763.  
PF 18-MAR-1998; 98US-0040963.  
PR (JACO/) JACOBS K.  
PA (MCCO/) MCCOY J M.  
PA (LAVALL) LAVALLIE E R.  
PA (COLL/) COLLINS-RACIE L A.  
PA (EVAN/) EVANS C.  
PA (MERB/) MERBERG D.  
PA (TREAC/) TREACY M.  
PA (SPAN/) SPaulding V.  
XX Jacobs K, McCoy JM, Lavallie ER, Collins-Racie LA, Evans C,  
PI Merberg D, Treacy M, Spaulding V;  
XX WPI; 2002-582343/62.  
DR P-PSDB; ABPE1799.  
XX Novel secreted or transmembrane protein and polynucleotide encoding the  
PT protein, useful for diagnosis and treatment of neurological disorders,  
PT cancer, autoimmune diseases, bone disorders and lung or liver fibrosis  
PT  
XX  
PS Claim 37; Page 111-112; 284dp; English.  
XX The invention relates to human secreted or transmembrane protein (I),  
XX their fragments and is encoded by specific complementary deoxyribonucleic  
XX acid (cDNA) inserts (II), where the protein is substantially free from  
XX other mammalian proteins. (I) are useful for preventing, treating or  
XX ameliorating a medical condition, especially immunological treatment or  
XX prevention of tumors. (I) exhibits activity relating to angiogenesis,  
XX cytokine, cell proliferation, cell differentiation, anti-inflammation,  
XX stem cell growth factor activity and activin or inhibin-related  
XX activities. (I) can be used to manipulate stem cells in culture to give  
XX rise to neuroepithelial cells that can be used to augment or replace  
XX cells damaged by illness, autoimmune disease, accidental damage or  
XX genetic disorders. (I) induces the proliferation of neural cells and  
XX regeneration of nerve and brain tissue and is useful for the treatment of  
XX central and peripheral nervous system diseases and neuropathies, such as  
XX Alzheimer's, Parkinson's disease, Huntington's disease, amyotrophic  
XX lateral sclerosis. (I) is involved in chemotactic or chemokinetic  
XX activity, regulation of haematopoiesis and is useful for treating myeloid  
XX or lymphoid cell disorders, platelet disorders such as thrombocytopenia  
XX and for regeneration of bone, cartilage, tendon, ligament and/or nerve  
XX tissue growth and in tissue repair, healing of burns, incisions, ulcers,  
XX for treating osteoporosis, osteoarthritis, bone degenerative disorders or  
XX periodontal disease. (I) is also useful for gut protection or  
XX regeneration and treatment of lung or liver fibrosis, repetitive injury  
XX in various tissues, various immune deficiencies and disorders including  
XX severe combined immunodeficiency (SCID), bacterial or fungal infections,  
XX autoimmune disorders e.g. multiple sclerosis, rheumatoid arthritis,  
XX diabetes mellitus, myasthenia gravis, allergic reactions and conditions,  
XX such as asthma or other respiratory problems. (II) is useful to express  
XX recombinant protein, as markers for tissues in which the corresponding  
XX protein is preferentially expressed and in gene therapy. The present  
XX sequence is that of a polynucleotide of the invention.  
SO Sequence 2199 BP; 552 A; 511 C; 674 G; 462 T; 0 other;

## Alignment Scores:

Pred. No.: 8.18e-106 Length: 2199  
Score: 1910.00 Matches: 359  
Percent Similarity: 100.00% Conservative: 0  
Best Local Similarity: 100.00% Mismatches: 0  
Query Match: 100.00% Indels: 0  
DB: 24 Gaps: 0

US-09-972-758A-2 (1-359) x ABQ92015 (1-2199)

QY 1 MetAlaGluProPheLeuSerGluTyrGlnHisGlnProGlnThrSerAsnCySThrGly 20  
Db ATGCCCGAGAGCATCTTGTGACAAATATCAACACAGCCTCAAACTAGACACTGTACAGGT 766  
QY 21 AlaAlaAlaValGlnGluGluLeuAsnProGluIuAagProProGlyValAGluGluIuArgVal 40  
Db GCTGCTGCTGCTCCAGAAAGAGTGAACCTTGAGCCGCCCGGAGGCGAGAGAGCGGGT 826  
QY 41 ProGluGluAspSerArgTyrGlnSerArgAlaPheProGlnLeuGlyGlyYarPProGly 60  
Db CCCGAGAGAGACAGTAGAGTGGCAATCGAGAGCGTTCCCGCACTGGGTGGCTCCGGG 886  
QY 61 ProGluGlyGluGlySerLeuGluSerGlnProProProLeuGlnThrGlnAlaCyPro 80  
Db CCGAGAGGGGAGAGGAGCTGGAAATCCCAACCACTCCCTTSCAAGCCAGCGCTGTCCA 946  
QY 81 GluSerSerCySLeuArgGluGlyGlyIuArgGlyGlnAsnGlyAspAspSerSerAlaGly 100  
Db GAATCTAGCTGCTGCTGAGAGAGGCGAGAAAGGCGCAAAATGGGGAGCACTGCTGCTG 1006  
QY 947 GAATCTAGCTGCTGCTGAGAGAGGCGAGAAAGGCGCAAAATGGGGAGCACTGCTGCTG 1006  
QY 101 GlysAppPheProProProAlaGluValGluProThrProGluValAGluLeuAsnAlaGln 120  
Db GCGACCTTCCCGCCCGCGGCAAGTGAACCGACCGCCGAGCGGAGCTGCTCCGCCAG 1066  
QY 121 ProCyHisAspSerGluAlaSerLySLeuGlyAlaProAlaAlaGlyGlyGluGluGlu 140  
Db CCTGTGATGATCTCCGAGCGCAAGTAAGTTGGGGCTCTCGCGGAGGGGCGAAGAGAG 1126  
QY 141 TrpGlyGlnGlnGlnArgGlnLeuGlyLySlySlyVHISArgArgProSerLySlyS 160  
Db TGGGAGACAGCAGACAGACAGCTGGGGAGAAAGAAAAACATAGAGAGCGCCCTCCAGAG 1186  
QY 161 LysArgHisTrpLyProTyrTrpLySLeuThrTrpGluGluLySlySlySlySlySly 180  
Db AGCGCATGTGAACCGTACTACAGCTGACCTGGAGAGAGAAAGAAAGTTCCAGCAG 1246  
QY 181 LysGlnSerLeuArgAlaSerArgIleArgAlaGluMetPheAlaLySlyGlyInProVal 200  
Db AAACAGAGCTTCGAGCTTCAGAGATCCGAGCGGAGATGTTGCCAAGGCGCAGCCGCT 1306  
QY 201 AlaProTyrAsnThrThrGlnPheLeuMetAspAspHisAspGlnGluGluProAspLeu 220  
Db GCGCCCTTAAACACACAGCAGGATCTCTCATGATGATCATCAGACAGAGAGAGCGGATCTC 1366  
QY 221 LysThrGlyLeuTyrSerLySlyArgAlaAlaAlaLySlySerAspAspThrSerAspAsp 240  
Db AAACCGGCTGTACTCTCAAGCGGCGCGCCCAATCCGAGACACCGCATGACGAC 1426  
QY 1367 AAACCGGCTGTACTCTCAAGCGGCGCGCCCAATCCGAGACACCGCATGACGAC 1426  
QY 241 PheMetGluGluGlyGlyGluGluAspGlyGlySerAspGlyMetGlyGlyYarPProGlySer 260  
Db TTCAATGAGAAAGGGGGTGAAGAGATGGGGCGACGATGGGATGGAGGGAGCGGCAGC 1486  
QY 1427 TTCAATGAGAAAGGGGGTGAAGAGATGGGGCGACGATGGGATGGAGGGAGCGGCAGC 1486  
QY 261 GluPheLeuGlnArgAspPheSerGluTyrGlyArgTyrHisThrGluSerLeuGln 280  
Db GAGTTTCTGACAGCGGAGCTTCTCGAGACGTCAGAGCGGTACCAACGAGAGAGCGCTGACG 1546  
QY 1487 GAGTTTCTGACAGCGGAGCTTCTCGAGACGTCAGAGCGGTACCAACGAGAGAGCGCTGACG 1546  
QY 281 AsnMetSerLySlyGlnLeuMetLysGlyTyrLeuGluLeuGluLySlySerLeuArg 300  
Db AACATGAGCAGACAGAGCTCATCAAGAGATCTGGAATCGAAGAGAGCTCTCGCC 1606  
QY 1547 AACATGAGCAGACAGAGCTCATCAAGAGATCTGGAATCGAAGAGAGCTCTCGCC 1606  
QY 301 MetGluAspGluAspAsnArgLeuArgLeuGluSerLySlyArgLeuGlyGlyYarPProGly 320  
Db ATGAGAGAGAGAGAGAGCGGCTGGGCTGGAGAGCAGAGCGGCTGGGTGGCCAGACGCG 1666  
QY 1607 ATGAGAGAGAGAGAGAGCGGCTGGGCTGGAGAGCAGAGCGGCTGGGTGGCCAGACGCG 1666  
QY 321 ArgValArgGluLeuGluGluLeuAspArgLeuArgAlaGluAsnLeuGlnLeuLeu 340  
Db CGTGTGCGGAGACTGAGAGCTGAGCTGAGCGCGCTGCGCGCGAGAAACCTCCAGCTGCTG 1726  
QY 1667 CGTGTGCGGAGACTGAGAGCTGAGCTGAGCGCGCTGCGCGCGAGAAACCTCCAGCTGCTG 1726  
QY 341 ThrGluAsnGluLeuHisArgGlnGlnArgAlaProLeuSerLySlyPheGlyYarP 359  
Db ACCGAGAACGAACTGACCGGAGAGAGAGCGCGCTTTCACAGTTTGGAGAC 1783  
QY 1727 ACCGAGAACGAACTGACCGGAGAGAGAGCGCGCTTTCACAGTTTGGAGAC 1783

RESULT 3





```

RESULT 4
ABN84013
ID ABN84013 standard; cDNA; 1080 BP.
XX
AC ABN84013;
XX
DT 10-SEP-2002 (first entry)
XX
DE Human oestrogen downregulated gene EDG1 coding sequence.
XX
KM EDG1; oestrogen downregulated gene; tumour suppressor; human;
KM breast cancer; prostate cancer; testicular cancer; ovarian cancer;
KM uterine cancer; colon cancer; chromosome 17q; gene therapy; gene;
KM ss.
XX
OS Homo sapiens.
XX
FH Key Location/Qualifiers
FT CDS 1..1080
FT /tag= a
FT /product= "EDG1"
FT /transl_except= "(pos:460..462,aa:Arg)"
FT /transl_except= "(pos:511..513,aa:Thr)"
XX
XX
XX WO200228879-A1.
XX
XX 11-APR-2002.
XX
XX 05-OCT-2001; 2001WO-US31300.
XX
XX 05-OCT-2000; 2000US-238187P.
XX
XX (UYCA-) UNIV CASE WESTERN RESERVE.
XX
XX Montano M, Wiltman B;
XX
XX WPI; 2002-519107/55.
XX
XX P-PSDB; ABB76495.
XX
XX Polynucleotides encoding Estrogen Down-Regulated Gene 1 proteins,
XX useful for the prevention, diagnosis and treatment of e.g. breast
XX cancer, testicular cancer, prostate cancer, uterine cancer, cervical
XX cancer, ovarian cancer and colon cancer -
XX
XX Claim 1; Fig 1A-B; 52pp; English.
XX
XX The present sequence is the coding sequence for human oestrogen
XX downregulated gene 1 (EDG1), a tumour suppressor gene that is
XX downregulated by oestrogen in mammary epithelial cells. The gene
XX was identified by yeast two-hybrid screenings for oestrogen
XX receptor-interacting proteins in breast epithelial cells. It
XX was localised to chromosome arm 17q. EDG1 mRNA expression is
XX prevalent in normal mammary epithelial cells and in other human
XX hormone-responsive tissues such as the ovary, prostate and testis.
XX Expression is low in breast cancer epithelial cells. Oestradiol,
XX which induces breast cancer cell growth, has an inhibitory effect
XX on EDG1 mRNA expression in breast cancer cells. Hexamethylene
XX bis-acetamide, an inducer of differentiation and apoptosis,
XX upregulates EDG1 mRNA expression in breast cancer cells. The
XX invention provides EDG1 polynucleotides and polypeptides. In a
XX claimed method, a test sample from an individual suspected of
XX having, or known to have breast, testicular, prostate, uterine,
XX cervical, ovarian or colon cancer is assayed for EDG1 transcript
XX using a polynucleotide that is complementary to the present
XX sequence or by RT-PCR using a primer derived from the present
XX sequence. A decrease in the level of transcript compared to the
XX level in a test sample indicates that the test sample contains or
XX was derived from cancerous cells antibody. A claimed method for
XX decreasing the proliferation of breast, prostate, testicular,
XX ovarian, uterine, cervical or colon cancer cells involves increasing
XX EDG1 protein activity in the cells, either by contacting the cells
XX with EDG1 protein or its fragment or functional equivalent, or with
XX a nucleic acid encoding EDG1 protein, its fragment or functional

```

```

CC equivalent.
XX
SQ Sequence 1080 BP; 265 A; 296 C; 376 G; 143 T; 0 other;
XX
Alignment Scores:
Pred. No.: 1,19e-105 Length: 1080
Score: 1902.00 Matches: 357
Percent Similarity: 99.72% Conservative: 1
Best Local Similarity: 99.44% Mismatches: 1
Query Match: 99.58% Indels: 0
DB: 24 Gaps: 0
US-09-972-758a-2 (1-359) x ABN84013 (1-1080)
QY 1 MetAlaGluProPheLeuSerGluTyrGlnHisGlnProGlnThrSerAsnCyThrGly 20
Db 1 ATGCCCGAGCCATTCTTGTCAGAAATATCAACCGCTCAACTAGCACTGTACAGGT 60
QY 21 AlaAlaAlaValGlnGluGluLeuAsnProGluArgProGluYAlaGluGluArgVal 40
Db 61 GCTGCTGCTGTCAGAGAGAGCTGAACCTGAGCGCCCCCGAGCGGAGGAGCGGGTG 120
QY 41 ProGluGluAspSerArgTTrpGlnSerAlaPheProGlnLeuGlyYArgProGly 60
Db 121 CCCGAGGAGGACAGTAGGTGGCAATCGAGAGCGTCCCAATGGTGGTGGCGGCGG 180
QY 61 ProGluGluGluGluSerLeuGluSerGlnProProProLeuGlnThrGlnAlaCysPro 80
Db 181 CCGAGAGGGGAGAGGAGCTGGAATCCACCACTCCCTTGACACCCAGGCTGTCCA 240
QY 81 GluSerSerCysLeuAlaArgGluGluYAluYsglyGlnAanglyAspAspSerSerAlaGly 100
Db 241 GAATCTAGCTGCTGTAGAGAGGCGCAGAAAGGCGCAAGATGGGAGAGACTCGCTGGC 300
QY 101 GlyAspPheProProProProAlaGluValGluProThrProGluAlaGluLeuAlaGln 120
Db 301 GCGGACTTCCCGCCCGCCGAGAACTGGAAACCGAGCCCGAGGCGGAGTGTCTCCCGCAG 360
QY 121 ProCysHisAspSerSerGluAlaSerLeuGluYAlaProAlaAlaGluYglYglGluGlu 140
Db 361 CTTTTCATGACTCCGAGGCGCAGTAGTGGGGCTCTCGCGGAGGGGCGCAAGAGGAG 420
QY 141 TrpGlyGlnGlnGlnArgGlnLeuGlyYslySlyShiSarArgArgProSerSlySly 160
Db 421 TGGGACACGACGACAGACAGACAGCTGGGGAGAAACAACTAAGACGCCCCGTCGAAGAAG 480
QY 161 LysArgHisTrpPLeuProTyrTyrLysLeuThrTrpGluGluYslyPhePheArgL 180
Db 481 AAGCGGCAATGGAAACCGTACTACAGCTGAACCTGGGAAGAAAGTTCCGACGAG 540
QY 181 LysGlnSerLeuAlaArgAlaSerArgLLeArgAlaGluMetPheAlaYslyGlnProVal 200
Db 541 AAACGAGACCTTCGAGCTTCAAGATCCGAGCCGAGGTTCGCAAGGGCCAGCGCGTGC 600
QY 201 AlaProTyrAsnThrThnGlnPheLeuMetAspAspHisAspGlnGluProAspLeu 220
Db 601 GCGCCCTTACACCAACGACGAGTCTCTCATGATGATCAACGACAGAGAGAGCGGATCTC 660
QY 221 LysThrGlyLeuTyrSerLysArgAlaAlaAlaYsSerAspAspThrSerAspAsp 240
Db 661 AAAACCGGCTGTACTTCCAAAGGGGCGCCCGCCCAATCCGACGACACGACGATGACGAC 720
QY 241 PheMetGluGluGluYglYglYglYsAspGlySerAspGlyYmeGlyYAspGlySer 260
Db 721 TTCAAGAGAAAGAGGGGTGAGAGATGGGGCACGATGGAGTGGAGGGAGCGGACGC 780
QY 261 GluPheLeuGlnArgAspPheSerGlnThrTyrGluArgYTrHisThrGlnSerLeuGln 280
Db 781 GAGTTTCTCCAGCGGAGACTTCTCGAGACGTHACAGCGGTACACACGAGAGAGCTCGAG 840
QY 281 AsnMetSerLysGlnGluLeuLysGlnTyrLeuGluLeuGluYslyCysLeuSerArg 300
Db 841 AACATGACCAAGCAGAGACTCATCAAGAGATGCTGGAATCGAAGAAAGTGCCTCTCGCGC 900

```

QY 301 MetGLuAspGluAsnAsnArgLeuArgLeuGluSerLysArgLeuGlyIAspAspAla 320  
 Db 901 ATGAGAGACGAGAACACCGGCTGCGGTGAGACAGCGGTGGTGGCGACGACGG 960  
 QY 321 ArgValArgGluLeuGluLeuGluLeuAspArgLeuArgAlaGluAsnLeuGlnLeu 340  
 Db 961 CGTGTCCGGAGCTGAGCTGAGCTGACCGGCTGCGCCGACGAACCTCCGCTGCTG 1020  
 QY 341 ThrGluAsnGluLeuHisArgGlnGlnArgAlaProLeuSerLysPheGlyAsp 359  
 Db 1021 ACCGAGACGAACTGACCGGCGAGAGAGCGCGCTTTCGAGTTTGAGAC 1077  
 RESULT 5  
 ABL90067/C  
 ID ABL90067 standard; cDNA; 1083 BP.  
 AC ABL90067;  
 XX  
 XX 24-MAY-2002 (first entry)  
 DE Human polynucleotide SEQ ID NO 629.  
 XX  
 XX Cytostatic; immunosuppressive; nootropic; neuroprotective; antiviral;  
 KW antidiabetic; hepatotropic; antidiabetic; antiinflammatory; antitumor;  
 KW vulnary; anticonvulsant; antibacterial; antifungal; antiparasitic;  
 KW cardiac; gene therapy; cancer; immune disorder; cardiovascular disorder;  
 KW neurological disease; infection; human; secreted protein; gene; ss.  
 OS Homo sapiens.  
 XX  
 XX MO200190304-A2.  
 XX  
 XX 29-NOV-2001.  
 XX  
 XX 18-MAY-2001; 2001MO-US16450.  
 XX  
 XX 19-MAY-2000; 2000US-205515P.  
 XX  
 XX (HUMA-) HUMAN GENOME SCI INC.  
 PA Birse CE, Rosen CA;  
 PI  
 XX  
 XX WPI; 2002-122018/16.  
 DR P-PSDB; ABB89658.  
 XX  
 XX Novel 1405 isolated polypeptides, useful for diagnosis, treatment and  
 PT prevention of neural, immune system, muscular, reproductive,  
 PT gastrointestinal, pulmonary, cardiovascular, renal and proliferative  
 PT disorders -  
 XX  
 XX Claim 4; SEQ ID NO 629; 2081bp + Sequence Listing; English.  
 XX  
 XX The invention relates to novel genes (ABL89449-ABL90853) and proteins  
 CC (ABB89040-ABB90444) useful for preventing, treating or ameliorating  
 CC medical conditions e.g. by protein or gene therapy. The genes are  
 CC isolated from a range of human tissues disclosed in the specification.  
 CC The nucleic acids, proteins, antibodies and (ant)agonists are useful  
 CC in the diagnosis, treatment and prevention of: (a) cancer, e.g. breast  
 CC and ovarian cancer and other cancers of the adrenal gland, bone, bone  
 CC marrow, breast, gastrointestinal tract, liver, lung, or urogenital;  
 CC (b) immune disorders e.g. Addison's disease, allergies, autoimmune  
 CC haemolytic anaemia, autoimmune thyroiditis, diabetes mellitus, Crohn's  
 CC disease, multiple sclerosis, rheumatoid arthritis and ulcerative  
 CC colitis; (c) cardiovascular disorders such as myocardial ischemias;  
 CC (d) wound healing; (e) neurological diseases e.g. cerebral anoxia and  
 CC epilepsy; and (f) infectious diseases such as viral, bacterial, fungal  
 CC and parasitic infections.  
 CC Note: The sequence data for this patent did not form part of the  
 CC printed specification, but was obtained in electronic format directly  
 CC from WIPO at ftp.wipo.int/pub/published\_pct\_sequences.  
 CC  
 XX Sequence 1083 BP; 301 A; 290 C; 199 G; 286 T; 7 other;

Alignment Scores:  
 Pred. No.: 1,43e-35 Length: 1083  
 Score: 733.00 Matches: 146  
 Percent Similarity: 98.654 Conservative: 0  
 Best Local Similarity: 98.658 Mismatches: 2  
 Query Match: 38.388 Indels: 1  
 DB: 24 Gaps: 0  
 US-09-972-758A-2 (1-359) x ABL90067 (1-1083)  
 QY 212 AspHisAspGlnGluGluProAspLeuLysThrGlyLeuTyrSerLysArgAlaAla 231  
 Db 1081 GATCAGACGACGAGAGAGCGGACCTCAAAACCGGCTTACTCCAAACGCGCCGCG- 1023  
 QY 232 LysSerAspAspThrSerAspAspPheMetGluGluGlyIAspGluLeuArgGly 251  
 Db 1022 AAATCCGACGACACCGCATGACCTTCATGAGAGAGGGGTGAGAGATGGGGGC 963  
 QY 252 SerAspGlyMetGlyIAspGlySerGluPheLeuGlnArgAspPheSerGluThrTyr 271  
 Db 962 AGCATGGGATGGAGGGGAGCGGACGAGCTTCTGACGGGACTTCTCGAGACGTAC 903  
 QY 272 GluArgTyrHisThrGluSerLeuGlnAsnMetSerLysGlnGluLeuLysGluTyr 291  
 Db 902 GAGCGGTACACACGAGAGCGCTGCAGAACATGAGACGACGAGAGCTCATAGAGGTAC 843  
 QY 292 LeuGluLeuGluLysCysLeuSerArgMetGluAspGluAsnAsnArgLeuArgLeu 311  
 Db 842 CTGGAATCGAAGAGTGCTCTTCGCGCATGAGAGACGAGAACACCGGCTGCGGAG 783  
 QY 312 SerLysArgLeuGlyIAspAspAlaArgValArgGluLeuGluLeuAspArg 331  
 Db 782 AGCAAGCGGCTGGTGGACGACGACGCGGCTGCGGAGACTGAGCTGAGACCGG 723  
 QY 332 LeuArgAlaGluAsnLeuGlnLeuLeuThrGluAsnGlnLeuHisArgGlnGlnGlu 351  
 Db 722 CTGCGGCGCGAGAACTCCAGCTGCTGACCGAAGCAACTCACCGGACGAGCGA 663  
 QY 352 AlaProLeuSerLysPheGlyAsp 359  
 Db 662 GCGCGGCTTCCAAAGTTTGGAGAC 639  
 RESULT 6  
 ID AAV86036  
 XX AAV86036 standard; cDNA; 523 BP.  
 AC AAV86036;  
 XX  
 XX 27-APR-1999 (first entry)  
 DE  
 XX  
 XX EST clone B227.  
 XX  
 XX Expressed sequence tag; secreted protein; hematopoiesis regulator;  
 KW tissue growth; activin; inhibin; tumour invasion suppressor; EST; human;  
 KW chemotaxis; chemokinesis; hemostasis; gene therapy; thrombolysis;  
 KW receptor; ligand; anti-inflammatory; tumour inhibitor; de.  
 OS Homo sapiens.  
 XX  
 XX MO9845435-A2.  
 PN  
 XX  
 XX 15-OCT-1998.  
 PD  
 XX  
 XX 10-APR-1998; 98MO-US06954.  
 PF  
 XX  
 XX 10-APR-1997; 97US-0835913.  
 PR  
 XX  
 XX (GENY ) GENETICS INST INC.  
 PA  
 XX Agostino M, Jacobs K, Lavallie ER, McCoy JM, Merberg D;  
 PI Racine LA, Spaulding V, Treacy M;  
 XX

DR WPI, 1999-070076/06.

XX New polynucleotides encoding human secreted proteins - derived from  
PT e.g. human blood, kidney, foetal lung, placenta, testes, brain,  
PT ovary, pituitary, retina and colon cDNA libraries

XX Claim 1; Page 97; 633pp; English.

XX This sequence represents an expressed sequence tag (EST), and is a  
CC polynucleotide of the invention. The polynucleotides of the invention are  
CC all secreted EST sequences isolated from a variety of human tissue  
CC sources. The EST sequences and proteins encoded by them are predicted to  
CC have useful biological activities which would make them suitable for  
CC treating, preventing or ameliorating medical conditions in humans and  
CC animals, although no supporting data is given. Suggested activities  
CC include nutritional activity, immune stimulating or suppressing activity,  
CC haematopoiesis regulating activity, tissue growth activity,  
CC activin/inhibin activity, chemotactic/chemokinetic activity, haemostatic  
CC and thrombolytic activity, receptor/ligand activity, anti-inflammatory  
CC activity, cadherin/tumour invasion suppressor activity, tumour inhibition  
CC activity. The EST sequences are also stated to be useful for gene  
CC therapy.

XX Sequence 523 BP; 112 A; 149 C; 153 G; 108 T; 1 other;

Alignment Scores:

Pred. No.:	Length:	Matches:	Conservative:
Score:	9,34e-35	523	111
Percent Similarity:	714.00		
Best Local Similarity:	99.26%		3
Query Match:	97.04%		1
	37.38%		0
		Gaps:	0

US-09-972-758a-2 (1-359) x AAV86036 (1-523)

```

QY 1 MetAlaGluProPheLeuSerGluTyrGlnHisGlnProGlnThrSerAncCysThrsly 20
DB 104 ATGGCCGAGCCATCTTCTTGACAAATATCAACACCACTCAACTGACAACTGACAGT 163
QY 21 A1A1A1A1A1A1G1G1G1G1G1G1G1G1G1G1G1G1G1G1G1G1G1G1G1G1G1G1 40
DB 164 GCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 223
QY 41 ProGluGluAAspSerArgTrpGlnSerArgAlaPheProGluLeuGlyGlyArgProGly 60
DB 224 CCGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 283
QY 61 ProGluGlyGluGlySerLeuGluSerGlnProProProLeuGlnThrGlnAlaCysPro 80
DB 284 CCGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 343
QY 81 GluSerSerCysLeuArgGluGlyGlyGlyGlyGlyGlyGlyGlyGlyGlyGlyGlyGly 100
DB 344 GAATCTAGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 403
QY 101 G1AAspPheProProProAlaGluValGluProThrProGluAlaGluLeuLeuAlaGln 120
DB 404 GGGGAGCTTCCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 463
QY 121 ProCysHisAspSerGluAlaSerGlyLeuGlyAlaPalaAla 135
DB 464 CTTGTCTCAAGACTCCGAGGCCAGTAAGATGGGTGCGACGCTCT 508

```

RESULT 7

ABX54966 ID ABX54966 standard; cDNA; 414 BP.

AC ABX54966;

XX 26-FEB-2003 (first entry)

XX Bovine EST associated with lactation/muscle/fat deposition #4895.

KW Bovine; ss; EST; expressed sequence tag; lactation; LMPD;  
KW muscle deposition; fat deposition; genome mapping; gene identification;  
KW gene analysis; cattle breeding.

OS Bos Taurus.

XX US2002137160-A1.

XX 26-SEP-2002.

XX 26-OCT-2001; 2001US-0983965.

XX 17-DEC-1998; 98US-113678P.

XX 15-DEC-1999; 99US-0465231.

XX (BYAT/) BYATT J C.

XX (MATH/) MATHIALAGAN N.

XX (TAON/) TAO N.

XX (WARR/) WARREN W C.

XX Byatt JC, Mathialagan N, Tao N, Warren WC;

XX WPI, 2003-102386/09.

Claim 2; SEQ ID No 4895; 38pp; English.

The invention relates to a purified nucleic acid molecule associated with  
lactation or muscle and fat deposition (designated LMPD), derived  
from cattle, and the LMPD nucleic acid can specifically hybridise to a  
second nucleic acid molecule comprising any of 5912 nucleotide  
sequences, appearing as ABX50072-ABX5983, or complements of them.  
Also included are: (1) a transformed cell having a nucleic acid  
comprising an LMPD nucleic acid linked to a promoter and a 3' non-  
translated sequence that functions in the cell to cause termination of  
transcription and addition of polyadenylated ribonucleotides to a 3' end  
of the mRNA molecule; and (2) determining a level or pattern of a  
marker nucleic acid in a bovine cell or tissue comprising: (a) incubating a marker  
nucleic acid (comprising any of the 5912 nucleic acid sequences or its  
complement or fragment) with a complementary nucleic acid molecule  
obtained from the bovine cell or tissue, where hybridisation between the  
marker nucleic acid and the complementary nucleic acid permits the  
detection of the molecule; and (b) detecting the level or pattern of the  
complementary nucleic acid, where the detection of the complementary  
nucleic acid is predictive of the level or pattern of the molecule.  
The LMPD nucleic acid is used for determining a level or pattern  
of a molecule in a bovine cell or tissue. It is useful for genome  
mapping, gene identification and analysis, cattle breeding, preparation  
of constructs for use in cattle gene expression, or for genetically  
improving cattle. The present sequence is one of the 5912 bovine  
CC LMPD EST (expressed sequence tag) nucleic acids.  
CC Note: The present sequence was not shown in the specification but  
CC was obtained in electronic format from the USPTO web site:  
CC [segdata.uspto.gov/sequence.html?docID=20020137160](http://segdata.uspto.gov/sequence.html?docID=20020137160).

XX Sequence 414 BP; 117 A; 106 C; 130 G; 61 T; 0 other;

Alignment Scores:

Pred. No.:	Length:	Matches:	Conservative:
Score:	1,76e-33	414	130
Percent Similarity:	691.00		
Best Local Similarity:	95.62%		1
Query Match:	94.89%		6
	36.18%		0
		Gaps:	0

US-09-972-758a-2 (1-359) x ABX54966 (1-414)

```

QY 155 ArgArgProSerGlyGlyGlyGlyGlyGlyGlyGlyGlyGlyGlyGlyGlyGlyGlyGly 174
DB 2 AGACGCCCTCCAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 61

```

```

QY 175 LysLysLysPheAspGluLysGlnSerLeuArgAlaSerArgIleArgAlaGluMetPhe 194
DB 62 AAGAAAAAGATCGATGAAAAACAGACCTGCGAGCTTCGAGGATTCGAGCCGAGATGTC 121
QY 195 AAlaLysGlyGlnProValAlaProTyrAerThrThGlnPheLeuMetAspPheAsp 214
DB 122 GCCAAGGGCCAGCAGCTTCTCCCTAATACACACACAGCTTCTCTATGATGACCCAGAC 181
QY 215 GlnGlnGlnProAspPheLysThrGlyLeuTyrSerLysArgAlaIleAlaLysSerAsp 234
DB 182 CAGGAGGAGCCGAGCTTTAAACCGGCTCTTATCCCAACGGGCGCTCCCAATCCGAC 241
QY 235 AspThrSerAspAspPheMetGluGluGlyGlyGluGluAspGlyLysSerAspGly 254
DB 242 GACACACGAGTATGAGATCTTTATGAAAGAGCGGCGAGAGAGATGCGGCGAGCGG 301
QY 255 MetGlyGlyAspGlySerGluPheLeuGlnArgAspPheSerGluThrTyrGluArgTyr 274
DB 302 ATGGGAGGAGACGCGCGAGCTTCTGCAAGCGGAGCTTCTCGGAGAACTATGAGCGGTAC 361
QY 275 HisThrGluSerLeuGlnAsnMetSerLysGlnGluLeuIleLeuGluTyr 291
DB 362 CACGCGGAGAGCTGCGAAGCATGACAGACGAGCTCATCAAGAGTAC 412

RESULT 8
ABL63543
ID ABL63543 standard; DNA; 461 BP.
XX ABL63543;
XX
XX 15-MAY-2002 (first entry)
DE Breast cancer related gene sequence SEQ ID NO:1880.
XX
XX Human; cancer; colon; breast; ovary; oesophagus; kidney; thyroid;
KW stomach; lung; prostate; pancreas; carcinoma; antitumor; cancerous;
KW cytostatic; gene therapy; antineoplastic; Wilms' tumour; adenocarcinoma;
KW gene; dr.
XX
XX Homo sapiens.
XX
XX WO200194629-A2.
XX
XX 13-DEC-2001.
XX
XX 30-MAY-2001; 2001MO-US10838.
XX
XX 05-JUN-2000; 2000US-209473P.
XX 05-JUN-2000; 2000US-209531P.
XX 18-SEP-2000; 2000US-23133P.
XX 18-SEP-2000; 2000US-231617P.
XX 20-SEP-2000; 2000US-234009P.
XX 20-SEP-2000; 2000US-234034P.
XX 20-SEP-2000; 2000US-234052P.
XX 22-SEP-2000; 2000US-234509P.
XX 22-SEP-2000; 2000US-234567P.
XX 25-SEP-2000; 2000US-234923P.
XX 25-SEP-2000; 2000US-234924P.
XX 25-SEP-2000; 2000US-235077P.
XX 25-SEP-2000; 2000US-235082P.
XX 25-SEP-2000; 2000US-235134P.
XX 25-SEP-2000; 2000US-235280P.
XX 26-SEP-2000; 2000US-235637P.
XX 26-SEP-2000; 2000US-235638P.
XX 27-SEP-2000; 2000US-235711P.
XX 27-SEP-2000; 2000US-235720P.
XX 27-SEP-2000; 2000US-235840P.
XX 27-SEP-2000; 2000US-235863P.
XX 28-SEP-2000; 2000US-236028P.
XX 28-SEP-2000; 2000US-236032P.
XX 28-SEP-2000; 2000US-236033P.
XX 28-SEP-2000; 2000US-236034P.

```

```

PR 28-SEP-2000; 2000US-236109P.
PR 28-SEP-2000; 2000US-236111P.
PR 29-SEP-2000; 2000US-236842P.
PR 29-SEP-2000; 2000US-236891P.
PR 02-OCT-2000; 2000US-237172P.
PR 02-OCT-2000; 2000US-237173P.
PR 02-OCT-2000; 2000US-237278P.
PR 02-OCT-2000; 2000US-237294P.
PR 02-OCT-2000; 2000US-237295P.
PR 02-OCT-2000; 2000US-237316P.
PR 03-OCT-2000; 2000US-237425P.
PR 03-OCT-2000; 2000US-237598P.
PR 03-OCT-2000; 2000US-237604P.
PR 03-OCT-2000; 2000US-237606P.
PR 03-OCT-2000; 2000US-237608P.
PR 01-NOV-2000; 2000US-244867P.
PR 01-NOV-2000; 2000US-245084P.

XX (AVAL-) AVALON PHARM.
XX
XX Young PE, Augustus M, Carter KC, Ebner R, Endress G, Horrigan S,
PI Soppet DR, Weaver Z;
XX WPI, 2002-188264/24.
XX
XX Screening for anti-neoplastic agent involves exposing cells to a
PT chemical agent to be tested for anti-neoplastic activity, and
PT determining a change in expression of a gene of a signature gene set -
XX
XX Claim 1, SEQ ID 1880, 44pp; English.
XX
XX The present invention describes a method (M1) for screening for an
CC anti-neoplastic agent. The method involves exposing cells to a chemical
CC agent to be tested for anti-neoplastic activity, determining a change in
CC expression of at least one gene (I) of a signature gene set, where (I)
CC comprises a sequence (S) selected from 8447 sequences (given in ABL61664
CC to ABL70110), or is at least 95% identical to (S), where a change in
CC expression is indicative of anti-neoplastic activity. (I) has cytostatic
CC activity and can be used in gene therapy. M1 can be used for screening
CC an anti-neoplastic agent, and can be used for producing a product which
CC is the data collected with respect to the anti-neoplastic agent as a
CC result of M1, and the data is sufficient to convey the chemical
CC structure and/or properties of the agent. M1 can be used in the
CC treatment of cancer such as colon, breast, stomach, lung, thyroid,
CC oesophageal, ovarian, kidney, prostate or pancreatic cancer,
CC adenocarcinoma, carcinoma, clear cell cancer, infiltrating ductal cancer,
CC infiltrating lobular cancer, squamous cell carcinoma, neuroendocrine
CC carcinoma, papillary carcinoma and Wilms' tumour.
XX
XX SQ Sequence 461 BP; 105 A; 134 C; 138 G; 80 T; 4 other;

Alignment Scores:
Pred. No.: 1 02e-29 Length: 461
Score: 629.00 Matches: 120
Percent Similarity: 93.85% Conservative: 2
Best Local Similarity: 92.31% Mismatches: 8
Query Match: 32.93% Indels: 1
DB: 24 Gaps: 0

US-09-972-758A-2 (1-359) x ABL63543 (1-461)
QY 1 MetAlaGluProPheLeuSerGluTyrGlnHisGlnProGlnThrSerAnCyThrGly 20
DB 72 ATGGCCGAGCCATTTCTTGTCAGAAATATCAACACGAGCTCAAACTATGCAACTGACAGGT 131
QY 21 AAlaAlaAlaValGlnGlnGluLeuAsnProGluArgProProGlyAlaGluGluArgVal 40
DB 132 GCTGCTGCTGTCAGGAAAGCTGAACCTGAGCGGCCCCCGAGCGGCGGCGGTG 191
QY 41 ProGluGluAspSerArgThrGlnSerArgAlaPheProGlnLeuGlyValArgProGly 60
DB 192 CCGAGAGAGACAGTATGATGTCGATGAGAGAGGCTTCCCAAGTGGGTGCGCGGG 251

```

QY 61. ProGluGlyGluGlySerLeuGluSerGlnProProProLeuGlnThrGlnAlaCysPro 80  
Db 252 CCGAGCGGGAGAGGAGCGCTGGAACTCCAACTCCCTTGAGAGCCAGGCGCTGTCCA 311  
QY 81 GluSerSerCysLeuArgGluGlyGlyGlyGlnAsnGlyAspAspSerSerAlaGly 100  
Db 312 GAATCTAGCTGCTGTGAGAGGGCGAGAAAGGCGCAAAATGGGAGCACTGCTCCGCTGAG 371  
QY 101 GlyAspPheProProProAlaGluValGluProThrProGluAlaGluLeuLeuAlaGln 120  
Db 372 GC-GACTTCCCGGCGCGGAGAAAGTGAGCCGACCCCGAGGCGCAAGCTTGACCA 430  
QY 121 ProCysHisAspSerGlnAlaSerIleu 130  
Db 431 CTTGTCTATGACTCCGAGGCGCACTAAGTTG 460  
RESULT 9  
AB061099  
ID AB061099 standard; cDNA; 1557 BP.  
XX  
AC AB061099;  
XX  
XX 26-FEB-2003 (first entry)  
XX  
DE Mouse putative protein #15 encoding sequence.  
XX  
XX Neuroprotective; immunomodulator; cancer;  
XX cytoskeletal; anti-inflammatory; gene therapy; nutritional supplement;  
XX wound; burn; ulcer; Alzheimer's disease; Huntington's disease;  
XX amyotrophic lateral sclerosis; autoimmune disorder; inflammation;  
XX vulnery; gene; ss.  
XX  
XX Mus musculus.  
XX  
XX WO200231111-A2.  
XX  
XX 18-APR-2002.  
XX  
XX 11-OCT-2001; 2001WO-US27760.  
XX  
XX 12-OCT-2000; 2000US-0687527.  
XX  
XX (HYSE-) HYSEQ INC.  
XX  
XX Tang YT, Liu C, Zhou P, Asundi V, Zhang J, Zhao QA, Ren F;  
XX Xue AJ, Yang Y, Wehrman T, Drmanac RT;  
XX  
XX WPI; 2002-426278/45.  
XX  
XX N-PSDB; ABP43855.  
XX  
XX New polypeptides and their encoded proteins, useful as nutritional  
XX sources or supplements, or in gene therapy, particularly for treating  
XX wounds, Alzheimer's disease, amyotrophic lateral sclerosis, cancer or  
XX inflammation  
XX  
XX Claim 1; SEQ ID # 312; 357bp + sequence listing; English.  
XX  
XX The invention relates to 446 newly isolated polynucleotide sequences.  
XX The activity of polynucleotides of the invention may be described as,  
XX vlnary, neuroprotective, immunomodulator, cytoskeletal and  
XX anti-inflammatory. Compositions comprising nucleic acids of the invention  
XX are useful for treating a mammalian subject, or as nutritional sources or  
XX supplements. These are useful in gene therapy, particularly for treating  
XX wounds, burns or ulcers, Alzheimer's disease, Huntington's disease,  
XX amyotrophic lateral sclerosis, autoimmune disorders, cancer or  
XX inflammation. The nucleic acids and polypeptides are also useful in  
XX diagnostic and research methods. The sequences given in records  
XX AB060788-AB061233 represent polynucleotides of the invention.  
XX NOTE: The sequence data for this patent did not form part of the printed  
XX specification, but was obtained in electronic format directly from Wipo  
XX at ftp.wipo.int/pub/published\_pct\_sequences.  
XX  
XX Sequence 1557 BP; 355 A; 426 C; 513 G; 263 T; 0 other;

Alignment Scores:  
Pred. No.: 3.6e-24 Length: 1557  
Score: 545.50 Matches: 131  
Percent Similarity: 57.28% Conservative: 42  
Best Local Similarity: 43.38% Mismatches: 99  
Query Match: 28.56% Gaps: 32  
DB: 24 Indels: 6  
US-09-972-758a-2 (1-359) x AB061099 (1-1557)  
QY 58 ArgProGlyGlyProGluGlyGlySerLeuGluSerGlnProProProLeuGlnThrGln 77  
Db 145 AGATTGGAACAGAGATGATGCA-CTCCGAAACAGACCGCTGTAATGACAGTAC 203  
QY 78 AlaCysProGluSerSerCysLeuArgGluGlyGlyGln-----AsnGly 94  
Db 204 CAGTGGCCCTG-GAGAGAGCCAAAGACTTGTGTGCTCCCGGGAGGCCCAACACCCCT 262  
QY 95 AspAspSerSerAlaGlyGlyAspPheProProAlaGluValGluProThrProGlu 114  
Db 263 GAGCGTCATGACTCTGTGTGTTCCCTGACACCGCGGATGAGACCTACAGAGCTG 322  
QY 115 AlaGluLeuLeuAlaGlnProCys-----HisAspSerGlnAlaSerIleu 130  
Db 323 GATGAAGATCTTGTGCGGCTGTGCTGCGTGGCTGGAACATGAGTGCCTCCGAGCC 382  
QY 131 GlyAlaProAlaAlaGlyGlyGluGluGluThrGlnGlnGlnAlaGluLeuGly 150  
Db 383 CAGACCCCGAGGGGCTGCTCAGCGGAG-----GCTGTGCTGCGCCG 424  
QY 151 LysLysHisArgArgArgProSerLysLysLysArgHisArgLysProTyrTyrLysLeu 170  
Db 425 AGAAACACCGCTCGCGGCGCATCGAAGCGCAAAAGCACTGCGACCTACCTGAGCTG 484  
QY 171 ThrTyrGluGluLysLysLysPheAspGluLysGlnSerLeuArgAlaSerArgIleArg 190  
Db 485 ACCTGGGCTGAGAAACAAACAGCGGATGAGAGCAGAGAGGCTCCCGGGTCCGC 544  
QY 191 AlaGluMetPheAlaLysGlyGlnProValAlaProTyrAsnThrThrGlnPheLeuMet 210  
Db 545 GAAAGAGATGTTGCCAAAGGCGAGCCCGTGGCCCTTCAACACCAACCCAGTTCCTGATG 604  
QY 211 AspAspHisAspGlnGluPro-----AspLeuLysThrGlyLeuTyrSerLysArg 228  
Db 605 AATGACAGGAGACCCGAGAGAGCCCAACTGATGTGCCCATGGATCTCCACCCAGGT 664  
QY 229 AlaAlaAlaLysSerAspAspThrSerAspAspAspPheMetGluGluGlyGluGlu 248  
Db 665 TCCAGTGGGAGAGT-----GAG 682  
QY 249 AspGlyGlySerAspGlyMetGlyLysAspGlySerGluPheLeuGlnArgAspPheSer 268  
Db 683 GCCGGGAGCAGTGAATGGCGGGCGGAGCGCAAGTGAATTTCCACGGAGAGACTTCTT 742  
QY 269 GluThrTyrGluArgTyrHisThrGluSerLeuGlnAsnMetSerLysGlnLeuLeu 288  
Db 743 GAGACTTACGAACGCTTCCACACCGAGACCTTCGAGCGCGGAGCAAGAGAGAGCTGCTG 802  
QY 289 LysGluTyrLeuGluLeuGluLysCysLeuSerArgMetGluAspGluAsnAsnArgLeu 308  
Db 803 CGAGACTACTCTGAGCTGAGAAAGCGGCTGTGCAAGCGGAGAGAGACTGAGAGCTG 862  
QY 309 ArgLeuGluSerLysArgLeuGlyLysAspAlaArg---ValArgGluLeuGluLeu 327  
Db 863 CAGCAGCTGACAGCGCTGACCCGCGAGAGTCTTCCGCTGAGTGAAGAGCTGCTGCC 922  
QY 328 GluLeuAspArgLeuArgAlaGluLeuLeuGlnLeuLeuThrGluAsnGluLeuHisArg 347  
Db 923 GAGGTTCAGAGGCTCGGAGCCGAAACAGCGGCTTCTGAGAGAAACCAAGATGTGGAAC 982  
QY 348 glnGln 349  
XXXXX

Db 983 CGAGAG 988

RESULT 10  
AAH99667/c  
ID AAH99667 standard; cDNA, 997 BP.  
XX  
XX  
XX AAH99667;  
DT 16-OCT-2001 (first entry)  
XX  
XX  
DE Human protein encoding cDNA sequence SEQ ID NO:502.  
KW Human; cancer; ulcer; HIV infection; human immunodeficiency virus;  
KW antiinflammatory; antirheumatic; antiarthritic; immunosuppressive;  
KW antibacterial; endocrine; cardiant; central nervous system; virucide;  
KW anti-HIV; fungicide; antimutagen; cardiovascular; antianaemic; anaemia;  
KW antiagregant; haemostatic; vulnery; antitumor; osteopathic; eczema;  
KW dermatological; antiallergic; antiasthmatic; antidiabetic; cytostatic;  
KW neuroprotective; antidepressant; nootropic; antiparkinsonian; infection;  
KW immunostimulant; gene therapy; antisense therapy; vaccine; inflammation;  
KW antianaphylactic; rheumatoid arthritis; septic shock; pancreatitis;  
KW cardiac dysfunction; neuropathology; cardiac anaphylaxis; autoimmunity;  
KW genetic disease; haematopoietic disorder; platelet disorder; asthma;  
KW thrombocytopaenia; osteoporosis; severe combined immunodeficiency;  
KW allergic rhinitis; diabetes; multiple sclerosis; depression;  
KW Alzheimer's disease; Parkinson's disease; neurodegenerative disorder;  
KW neurological disorder; se.  
XX  
XX Homo sapiens.  
XX  
XX MO200153455-A2.  
XX  
XX 26-JUL-2001.  
XX  
XX  
XX 22-DEC-2000; 2000MO-US35017.  
XX  
XX  
XX 23-DEC-1999; 99US-0471275.  
XX 21-JAN-2000; 2000US-0488725.  
XX 23-APR-2000; 2000US-0552317.  
XX  
XX  
XX (HYSE-) HYSEQ INC.  
XX  
XX  
XX Tang YT, Liu C, Drmanac RT;  
XX  
XX WPI; 2001-457603/49.  
XX  
XX P-PSDB; AAM253726.  
XX  
XX  
XX Isolated human polynucleotides encoding polypeptides, useful for the  
XX treatment and diagnosis of e.g. cancer, ulcers and HIV infection -  
XX  
XX  
XX Claim 1; Page 566-567; 1217bp; English.  
XX  
XX  
XX AAH99166 to AAH99904 encode the human proteins given in AAM25225 to  
XX CC AAM25963. The proteins can have activities based on the tissues and  
XX CC cells they are expressed in, such as: antiinflammatory; antirheumatic;  
XX CC antiarthritic; immunosuppressive; antibacterial; endocrine; cardiant;  
XX CC central nervous system; virucide; anti-HIV; fungicide; antimutagen;  
XX CC cardiovascular; antianaemic; antiagregant; haemostatic; vulnery;  
XX CC antitumor; osteopathic; dermatological; antiallergic; antiasthmatic;  
XX CC antidiabetic; cytostatic; neuroprotective; antidepressant; nootropic;  
XX CC antiparkinsonian; and immunostimulant. The proteins and polynucleotides  
XX CC encoding them can be used in gene therapy, antisense therapy and vaccine  
XX CC production. The proteins and polynucleotides are useful for screening for  
XX CC agents or antagonists of a protein and for the treatment and diagnosis  
XX CC of disorders associated with the activity of a protein e.g. inflammation,  
XX CC rheumatoid arthritis, septic shock, pancreatitis, cardiac dysfunction,  
XX CC neuropathology, cardiac anaphylaxis, viral, bacterial, HIV and fungal  
XX CC infections, autoimmunity, genetic diseases, haematopoietic disorders,  
XX CC anaemia, platelet disorders, thrombocytopaenia, wounds, burns, ulcers,  
XX CC osteoporosis, severe combined immunodeficiency, eczema, allergic  
XX CC rhinitis, asthma, diabetes, cancer, multiple sclerosis, depression,  
XX CC Alzheimer's disease, Parkinson's disease, neurodegenerative and  
XX CC neurological disorders.

XX	Sequence	997 BP, 184 A, 323 C, 266 G, 224 T, 0 other;
US	US-09-972-758A-2 (1-359) x AAV99667 (1-997)	
Alignment Scores:		
Pred. No.:	3,68e-10	997
Score:	308.50	Matches: 70
Percent Similarity:	60.26%	Conservative: 21
Best Local Similarity:	46.16%	Mismatches: 43
Query Match:	16.15%	Indels: 17
DB:	22	Gaps: 3
QY	202	ProTyAspThrThGluPheLeuMetAspAspHisAspGlnGluGluPro-----Asp 219
DB	980	CCCTTAACAACACACCCAGTTCTCTGATGATGACAGGAGCCGGAGAGACCACTTGAT 921
QY	220	LeuTySerThGluLeuTySerLysArgAlaAlaLysSerAspAspThrSerAsp 239
DB	920	GTGCCCCATGGGATCTCCACCCACAGTTCCAGTGGGAGAGT----- 879
QY	240	AspPheMetGluGluGlyGlyGluGluAspGlyGlySerAspGlyMetGlyGlyAspGly 259
DB	878	-----GAGCCCGGGACAGTGATGGCGGGCCGAGCCGAC 843
QY	260	SerGluPheLeuGluArgAspPheSerGluThrTyGluArgTyRHisThrGluSerLeu 279
DB	842	GGTAGTTCCAGCGAGAGACTTCTCTAGACTTACGAACTTCCACACCGAGAGCTG 783
QY	280	GlnAsMetSerLysGlnGluLeuLeuLeuLeuGluTyRLeuGluLeuGluLysCysLeuSer 299
DB	782	CAGGCGCGACGACAGAGAGAGCTGGTGAGACTACCTCGAGCTCGGAGAAAGCGGCTCG 723
QY	300	ArgMetGluAspGluAsnAsnArgLeuArgLeuGluSerLysArgLeuGlyGlyAspAsp 319
DB	722	CAGGCGGAGAGAGACTAGAGGCTGCAGACGCTGCAGCGGCTCACCGGCGACGACTCC 663
QY	320	AlaArg---ValArgGluLeuGluLeuGluLeuAspArgLeuArgAlaGluAsnLeuGln 338
DB	662	TGCGGCGAGGTGAGAGAGCTGGCTGCCGAGGTCCAGAGGCTCCGAGACCGAAACAGCGG 603
QY	339	LeuLeuThrGluAsnGluLeuHisIArgGlnGln 349
DB	602	CTTCGTGAGAGAACCAAGTGTGGAAACCGAGAG 570
RESULT 11		
AAV87316		
ID	AAV87316 standard; cDNA; 342 BP.	
XX		
AC	AAV87316;	
XX		
DT	27-APR-1999 (first entry)	
XX		
DE	EST clone BP783.	
KW	Expressed sequence tag; secreted protein; haematopoiesis regulator;	
KW	tissue growth; activin; inhibin; tumour invasion suppressor; EST; human;	
KW	chemotaxis; chemokinesis; haemostasis; gene therapy; thrombolysis;	
KW	receptor; ligand; anti-inflammatory; tumour inhibitor; ds.	
OS	Homo sapiens.	
XX		
PN	W09845435-A2.	
XX		
PD	15-OCT-1998.	
XX		
PF	10-APR-1998; 98WO-US06954.	
XX		
PR	10-APR-1997; 97US-0835913.	
XX		
PA	(GENEY ) GENETICS INST INC.	
TI	Apostolno M, Jacobs K, Lavallie ER, McCoy JM, Merberg D;	









XX 19-FEB-2001 (first entry)  
 DT Cat flea head and nerve cord (HNC) cDNA, SEQ ID NO:1712.  
 DE  
 XX  
 XX Cat flea; head and nerve cord nucleic acid; HNC; flea infestation;  
 KM vaccine; antiparasitic; therapeutic target; diagnosis; detection; ss.  
 XX  
 XX Ctenocephalides felis.  
 OS  
 XX MO200061621-A2.  
 PN  
 PD 19-OCT-2000.  
 XX  
 XX 07-APR-2000; 2000MO-US09437.  
 PF  
 XX 09-APR-1999; 99US-0128704.  
 PR  
 XX (HESK-) HESKA CORP.  
 PA  
 XX Brandt KS, Gaines PJ, Stinchcomb DT, Wisniewski N;  
 PI WPI; 2000-656323/63.  
 PS  
 XX  
 XX Claim 26; Page 818; 964pp; English.  
 PT  
 PT The invention relates to novel cat flea (Ctenocephalides felis) nucleic  
 acids which are expressed in hindgut and Malpighian tubule (HMT) tissue  
 or head and nerve cord (HNC) tissue. The invention also relates to the  
 encoded proteins. The invention additionally encompasses expression  
 constructs, recombinant viruses and recombinant cells comprising the  
 nucleic acids of the invention, recombinant production of the proteins,  
 antibodies against the proteins, a method of identifying inhibitors of  
 the proteins, and compositions comprising the inhibitors for  
 administration to an animal. The nucleic acids, and the proteins they  
 encode may be used in the prevention, treatment and diagnosis of diseases  
 associated with flea infestations. For example, the nucleic acids may be  
 used to produce an HMT or HNC protein according to standard recombinant  
 DNA methodology by inserting the nucleic acids into a host cell and  
 culturing the cell to express the protein. The HMT and HNC nucleic acids  
 may also be used as DNA probes in diagnostic assays (e.g., PCR) to detect  
 and quantitate the presence of cat flea or other homologous nucleic acid  
 sequences in samples. They may also be used to study the expression and  
 function of the proteins and their role in metabolism. The HMT and HNC  
 proteins may be used as antigens in the production of specific  
 antibodies, and in assays to identify modulators (agonists and  
 antagonists) of HMT and/or HNC protein expression and activity. The  
 anti-HMT/HNC protein antibodies and antagonists may also be used to  
 CC downregulate protein expression and activity. The antibodies may also be  
 CC used as diagnostic agents for detecting the presence of flea polypeptides  
 CC in samples (e.g., by enzyme linked immunosorbent assay (ELISA)). The  
 CC present sequence represents a cat flea HNC cDNA of the invention.  
 XX  
 XX  
 SQ Sequence 549 BP; 207 A; 100 C; 113 G; 126 T; 3 other;

Alignment Scores:

Pred. No.:	Length:	549
Score:	184.00	Matches: 66
Percent Similarity:	42.29%	Conservative: 30
Best Local Similarity:	29.07%	Mismatches: 53
Query Match:	9.63%	Indels: 12
DB:	21	Gaps: 18

US-09-972-758a-2 (1-359) x AAC95217 (1-549)

QY 34 ProGlyAlaGluGluArgValProGluGluAspSerArgTrpGlnSerArgAlaPhe---52  
 DB 33 CCAGCGNGGTGAAGAAATAGTG-----AGTAAATATCAATCAAAAGTGTITTTTA 80

QY 53 ---ProGlnLeuGlyAlaArgProGluGluGluGlySerLeuGluSerGlnPro 71  
 DB 81 TTAATACCTTGACATCAATAATAGTAGTAAATTTGAA---AATATTGAATCGGAAT 137  
 QY 72 Pro-----ProLeuGlnThrGlnAla 78  
 DB 138 CCAAAATTAATACAAACCGAGTACAGCGTCCAAATGTCTGCCCTTCATAACCAAGCGCT 197  
 QY 79 CysProGluSerSerCysLeuArgGluGluGlyGlnAsnGlyAspAspSerSer 98  
 DB 198 -----ACTGAAGCGCGCTTATCAGAGAGAAAAGGAATAGATGAT----- 239  
 QY 99 AlaGlyGlyAspPheProProProAlaGluValGluProThrProGluAlaGluLeu 118  
 DB 239 ----- 239  
 QY 119 AlaGlnProCysHisAspSerGlnAlaSerLeuGlyAlaProAlaAlaGlyGlyGlu 138  
 DB 240 -----AGAAAGAGAGCAACCAAGTGGT-----GCAACAAATGGGAAG 278  
 QY 139 GluGluTrpGlyGlnGlnArgGlnLeuGlyLeuGlySerHisArgArgProSer 158  
 DB 279 -----AGAAAACATAGACGAGTAAATGT 305  
 QY 159 LysLysLysArgHisTrpLysProTyrTyrLysLeuThrTrpGluGluLysLysPhe 178  
 DB 306 AAGAGAAAACCTAAT---AAACGTAACAATAAAGCCGTGACCAAGTAAGATGTC 362  
 QY 179 AspGluLysGlnSerLeuArgAlaSerArgIleArgAlaGluMetPheAlaLysGln 198  
 DB 363 ---CAAAAGAAATGATGAGAGTTCGTAGTGCCTGTCTAAATATTAAGCCATGGGCAAC 419  
 QY 199 ProValAlaProTyrAsnThrThrGlnPheLeuMetAspAspHisAspGlnGluPro 218  
 DB 420 ACATTAGTTCCTGTAAACGACCAATTTCTCATGGAAGATCATGAT----- 467  
 QY 219 AspLeuLysThrGlyLeuTyrSerLysArgAlaAlaAlaLysSerAspAspThr----- 236  
 DB 468 -----GTCTCACCAGGATTCATCTGCGAGCTCGACTCTACTTAGTGT 512  
 QY 237 ---SerAspAspAspPheMet 242  
 DB 513 CGTCTGAAGACACTCTAATG 533

RESULT 15  
 AAS79493  
 ID AAS79493 standard; cDNA; 1024 BP.  
 XX  
 AC AAS79493;  
 XX  
 DT 13-FEB-2002 (first entry)  
 XX  
 DE DNA encoding novel human diagnostic protein #15297.  
 XX  
 XX Human; chromosome mapping; gene mapping; gene therapy; forensic;  
 KM Food supplement; medical imaging; diagnostic; genetic disorder; ss.  
 XX  
 OS Homo sapiens.  
 XX  
 PN WO200175067-A2.  
 PD 11-OCT-2001.  
 XX  
 PF 30-MAR-2001; 2001MO-US08631.  
 XX  
 PR 31-MAR-2000; 2000US-0540217.  
 PR 23-AUG-2000; 2000US-0649167.  
 XX  
 PA (HYSE-) HYSEQ INC.  
 XX  
 XX Drmanac RT, Liu C, Tang YT;  
 PI WPI; 2001-639362/73.  
 DR





XX (UYCA-) UNIV CASE WESTERN RESERVE.  
 PA  
 XX  
 XX Montano M, Wiltman B;  
 PI  
 XX  
 DR WPI: 2002-519107/55.  
 DR P-PSDB; ABB76495.  
 XX  
 PT Polynucleotides encoding Estrogen Down-Regulated Gene 1 proteins,  
 PT useful for the prevention, diagnosis and treatment of e.g. breast  
 PT cancer, testicular cancer, prostate cancer, uterine cancer, cervical  
 PT cancer, ovarian cancer and colon cancer -  
 XX  
 PS Claim 1; Fig 1A-B; 52pp; English.  
 XX  
 CC The present sequence is the coding sequence for human oestrogen  
 CC downregulated gene 1 (EDG1), a tumour suppressor gene that is  
 CC downregulated by oestrogen in mammary epithelial cells. The gene  
 CC was identified by yeast two-hybrid screenings for oestrogen  
 CC receptor-interacting proteins in breast epithelial cells. It  
 CC was localised to chromosome arm 17q. EDG1 mRNA expression is  
 CC prevalent in normal mammary epithelial cells and in other human  
 CC hormone-responsive tissues such as the ovary, prostate and testis.  
 CC Expression is low in breast cancer epithelial cells. Oestradiol,  
 CC which induces breast cancer cell growth, has an inhibitory effect  
 CC on EDG1 mRNA expression in breast cancer cells. Hexamethylene  
 CC bis-acetamide, an inducer of differentiation and apoptosis,  
 CC upregulates EDG1 mRNA expression in breast cancer cells. The  
 CC invention provides EDG1 polynucleotides and polypeptides. In a  
 CC claimed method, a test sample from an individual suspected of  
 CC having, or known to have breast, testicular, prostate, uterine,  
 CC cervical, ovarian or colon cancer is assayed for EDG1 transcript  
 CC using a polynucleotide that is complementary to the present  
 CC sequence or by RT-PCR using a primer derived from the present  
 CC sequence. A decrease in the level of transcript compared to the  
 CC level in a test sample indicates that the test sample contains or  
 CC was derived from cancerous cells antibody. A claimed method for  
 CC decreasing the proliferation of breast, prostate, testicular,  
 CC ovarian, uterine, cervical or colon cancer cells involves increasing  
 CC EDG1 protein activity in the cells, either by contacting the cells  
 CC with EDG1 protein or its fragment or functional equivalent, or with  
 CC a nucleic acid encoding EDG1 protein, its fragment or functional  
 CC equivalent.  
 CC  
 XX  
 SQ Sequence 1080 BP; 265 A; 296 C; 376 G; 143 T; 0 other;  
 Query Match 100.0%; Score 1080; DB 24; Length 1080;  
 Best Local Similarity 100.0%; Pred. No. 1,1e-213;  
 Matches 1080; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 361 CCTTGTCAATGACTCCGAGGCCAGTAAGTTGGGGGCTCTGCCGACAGGGGCGAAGAGAG 420  
 DB 361 CCTTGTCAATGACTCCGAGGCCAGTAAGTTGGGGGCTCTGCCGACAGGGGCGAAGAGAG 420  
 QY 421 TGGGGAACAGCAGAGAGACAGCTGGGGAGAAAAAATTAAGAGACCCCTTCAAGAG 480  
 DB 421 TGGGGAACAGCAGAGAGACAGCTGGGGAGAAAAAATTAAGAGACCCCTTCAAGAG 480  
 QY 481 AAGCGCATTTGGAAACCGTACTACAACTGAATCTGGGAAGAGAAAAAGTTTCAGAG 540  
 DB 481 AAGCGCATTTGGAAACCGTACTACAACTGAATCTGGGAAGAGAAAAAGTTTCAGAG 540  
 QY 541 AAACAGAGCCTTGAGCTTCAAGATCCGAGCCGAGATTTGCCAAGGCCAGCCGATC 600  
 DB 541 AAACAGAGCCTTGAGCTTCAAGATCCGAGCCGAGATTTGCCAAGGCCAGCCGATC 600  
 QY 601 GCGCCTTAATACACACGAGCTTCTCATGATGATCAACAGACAGAGAGCCGATCTC 660  
 DB 601 GCGCCTTAATACACACGAGCTTCTCATGATGATCAACAGACAGAGAGCCGATCTC 660  
 QY 661 AAAACGGGCTGTACTCAAGGGGGGCGCGCAATCCGAGACACGCGATGACAG 720  
 DB 661 AAAACGGGCTGTACTCAAGGGGGGCGCGCAATCCGAGACACGCGATGACAG 720  
 QY 721 TTCAATGAAGAAAGGGGGTGAAGAGATGGGGGACGATGGAGTGGAGAGGAGCGAC 780  
 DB 721 TTCAATGAAGAAAGGGGGTGAAGAGATGGGGGACGATGGAGTGGAGAGGAGCGAC 780  
 QY 781 GAGTTTCTGACGAGGGACTTCTCGAGAGCTACAGCGGTACCAACGAGAGCCTGCAG 840  
 DB 781 GAGTTTCTGACGAGGGACTTCTCGAGAGCTACAGCGGTACCAACGAGAGCCTGCAG 840  
 QY 841 AACATGAGACAGAGAGAGCTCATCAAGAGTACTGGAACTGGAAGTGGCTTCGCGC 900  
 DB 841 AACATGAGACAGAGAGAGCTCATCAAGAGTACTGGAACTGGAAGTGGCTTCGCGC 900  
 QY 901 ATGAGAGACGAGAAACAAACGGGCTGCGGCTGAGAGACAAACGGGCTGGGTCGACAGCGG 960  
 DB 901 ATGAGAGACGAGAAACAAACGGGCTGCGGCTGAGAGACAAACGGGCTGGGTCGACAGCGG 960  
 QY 961 CGTGTGCGGAGCTGAGAGCTGAGAGCTGAGACCGGCTGCGGCGAGAACTTCCAGCTGCTG 1020  
 DB 961 CGTGTGCGGAGCTGAGAGCTGAGAGCTGAGACCGGCTGCGGCGAGAACTTCCAGCTGCTG 1020  
 QY 1021 ACCGAGAACGAACCTGACCGGAGACGAGAGCGGCGCTTCCAGTTTGAAGCTAG 1080  
 DB 1021 ACCGAGAACGAACCTGACCGGAGACGAGAGCGGCGCTTCCAGTTTGAAGCTAG 1080

RESULT 2  
 ID AAV82778  
 ID AAV82778 standard; cDNA; 2199 BP.  
 XX  
 AC AAV82778;  
 XX  
 DT 25-FEB-1999 (first entry)  
 XX  
 DE Clone bp783\_3 isolated from human foetal kidney cDNA library.  
 XX  
 KW Secreted protein; nutritional activity; immune stimulating; vaccine;  
 KW suppressing activity; haematopoiesis regulating activity;  
 KW tissue growth activity; activin; inhibin activity; chemokinesis;  
 KW chemokinetic activity; haemostasis; thrombolytic activity; receptor;  
 KW ligand; anti-inflammation; cadherin; tumour invasion suppressor;  
 KW tumour inhibition; gene therapy; ds.  
 XX  
 OS Homo sapiens.  
 XX  
 PN MO9842739-A2.  
 XX  
 PD 01-OCT-1998.  
 XX

20-MAR-1998; 98WO-US05653 .  
 19-MAR-1998; 98US-0044466 .  
 21-MAR-1997; 97US-0822167 .  
 (GEMY ) GENETICS INST INC .  
 Agostino MJ, Jacobs K, Lavallie ER, McCoy JM, Merberg D;  
 Racie LA, Spaulding V, Treacy M;  
 WPI, 1998-609890/51 .  
 P-PSDB; AAM85455 .  
 New polynucleotides encoding secreted human proteins - derived from  
 human foetal brain, adult brain, foetal kidney, placenta or adult  
 pineal gland cDNA libraries .  
 Claim 1; Page 66-67; 113pp; English .  
 The present sequence encodes a secreted protein. The polynucleotide and  
 secreted protein are predicted to have biological activities which would  
 make them suitable for treating, preventing or ameliorating medical  
 conditions in humans and animals, although no supporting data is given.  
 Suggested activities include nutritional activity, immune stimulating  
 (e.g. as vaccines) or suppressing activity, haematopoiesis regulating  
 activity, tissue growth activity, activin/inhibin activity,  
 chemotactic/chemokinetic activity, haemostatic and thrombolytic activity,  
 receptor/ligand activity, anti-inflammatory activity, cadherin/tumour  
 invasion suppressor activity, and tumour inhibition activity (no data is  
 given in the specification to support these activities) . The  
 polynucleotide is also stated to be useful for gene therapy .  
 Sequence 2199 BP; 552 A; 511 C; 674 G; 462 T; 0 other;

Query Match	99.7%	Score 1076.8	DB 19	length 2199
Best Local Similarity	99.8%	Pred No. 5.9e-213		
Best Local, Conservative	0	Mismatches 2	Indels 0	Gaps 0

QY	1	TTGGCCGAGCAATTTTGTCAAGATATCAACACAGGCTCAAACTAGCAATCTGTCAAGT	60
Db	707	ATGGCCGAGCCATTTTGTCAAGATATCAACACAGGCTCAAACTAGCAATCTGTCAAGT	766
QY	61	GCTGCTGCTGTCCAGAGAAAGCTGAACCTTGAGCGCCCCCAAGCGCGAGAGCGGGTGTG	120
Db	767	GCTGCTGCTGTCCAGAGAAAGCTGAACCTTGAGCGCCCCCAAGCGCGAGAGCGGGTGTG	826
QY	121	CCCGAGGAGGACAGTAAAGTGTGGCAATCCAGAGCGCTTCCCCCAATTGGGTGGCGGTCCGGGG	180
Db	827	CCCGAGGAGGACAGTAAAGTGTGGCAATCCAGAGCGCTTCCCCCAATTGGGTGGCGGTCCGGGG	886
QY	181	CCGAGAGGGGAAAGGAGGCTGTGAAATCCCAACCACTCTCTTTCAGAACCCAGGCTGTGCCA	240
Db	887	CCGAGAGGGGAAAGGAGGCTGTGAAATCCCAACCACTCTCTTTCAGAACCCAGGCTGTGCCA	946
QY	241	GAATCTACTGTCTGAGAGAGGGGCGAGAAAGGGCCAGAAATGGGGAGAGACTGTGTCCGTGGC	300
Db	947	GAATCTACTGTCTGAGAGAGGGGCGAGAAAGGGCCAGAAATGGGGAGAGACTGTGTCCGTGGC	1006
QY	301	GCGCAGTTCCCGCCCGCCAGCAAGATGTGAAACCGACGCCCCGAGGCGGAAGTGTGTGCCAG	360
Db	1007	GCGCAGTTCCCGCCCGCCAGCAAGATGTGAAACCGACGCCCCGAGGCGGAAGTGTGTGCCAG	1066
QY	361	CCTTGTCAATGACTCCGAGGCCAGTAAAGTTGGGGGGCTCTCTCCCGCAGGGGGCGAANAAGAG	420
Db	1067	CCTTGTCAATGACTCCGAGGCCAGTAAAGTTGGGGGGCTCTCTCCCGCAGGGGGCGAANAAGAG	1126
QY	421	TGGGAGACAGCAGCAGACAGACAGCTGGGGGAGAAAAAACAATAGAGACGCGCCGTCCAAGAG	480
Db	1127	TGGGAGACAGCAGCAGACAGACAGCTGGGGGAGAAAAAACAATAGAGACGCGCCGTCCAAGAG	1186
QY	481	AAGCGGCAATTGGAAACCGTAACTACAGCTGAACTGGGAAAGAGAAAGAAAAAGTTTGCACGAG	540
Db	1187	AAGCGGCAATTGGAAACCGTAACTACAGCTGAACTGGGAAAGAGAAAGAAAAAGTTTGCACGAG	1246

QY	541	AAACAGAGCCCTGAGCTTCAAGAGTCCGAGCCGAGATGTTCCGCAAGGGCCAGCCGCTC	600
Db	1247	AAACAGAGCCCTGAGCTTCAAGAGTCCGAGCCGAGATGTTCCGCAAGGGCCAGCCGCTC	1306
QY	601	GCGCCCTATAACACACAGCAGTTCCTCATGATGATATCACACAGAGAGAGCCGGATCTC	660
Db	1307	GCGCCCTATAACACAGCAGTTCCTCATGATGATATCACACAGAGAGAGCCGGATCTC	1366
QY	661	AAAAACGGGCTGTATCTCCAAAGCGGGCCGCGCCAAATCCGACGACACACAGGATGACGAC	720
Db	1367	AAAAACGGGCTGTATCTCCAAAGCGGGCCGCGCCAAATCCGACGACACACAGGATGACGAC	1426
QY	721	TTCTATGGAAGAAGGGGGGTGAGAGAGATGGGGGGCAGACGATGGGATGGGAAGGGGACCGGAGC	780
Db	1427	TTCTATGGAAGAAGGGGGGTGAGAGAGATGGGGGGCAGACGATGGGATGGGAAGGGGACCGGAGC	1486
QY	781	GAGTTTCTGACGCGGAGCTTCTCGAGACGTAACGAGGGTATCCACACAGGAGAGCCTTCGAC	840
Db	1487	GAGTTTCTGACGCGGAGCTTCTCGAGACGTAACGAGGGTATCCACACAGGAGAGCCTTCGAC	1546
QY	841	AACATGAGCAAGCAGAGCTCATTCMAAGAGTACTCTGGAACCTGGAGACTGGAGAGTGCCTTCTCGC	900
Db	1547	AACATGAGCAAGCAGAGCTCATTCMAAGAGTACTCTGGAACCTGGAGACTGGAGAGTGCCTTCTCGC	1606
QY	901	ATGAGGAGACGAGAACACCCGGCTTCGCGCTGAGAGCAGCGCGCTGGGTGGCGACGACCGC	960
Db	1607	ATGAGGAGACGAGAACACCCGGCTTCGCGCTGAGAGCAGCGCGCTGGGTGGCGACGACCGC	1666
QY	961	CGTGTGGGGAGCGCTGAGAGCTTGAGACCGGCTGGCGCCCGAGAACCTTCAGCTCTG	1020
Db	1667	CGTGTGGGGAGCGCTGAGAGCTTGAGACCGGCTGGCGCCCGAGAACCTTCAGCTCTG	1726
QY	1021	ACCGAGAACGAACTGCACCGGACGACGAGACGAGCGCGCTTTCCAAGTTTGGAGACTAG	1080
Db	1727	ACCGAGAACGAACTGCACCGGACGACGAGACGAGCGCGCTTTCCAAGTTTGGAGACTAG	1786

RESULT 3	ABQ92015	ABQ92015 standard; cDNA; 2199 BP.
XX	ABQ92015	
AC		
XX		
DT	04-OCT-2002 (first entry)	
XX		
DE	Human polynucleotide SEQ ID NO 12.	
XX		
KW	Human; cytostatic; antirheumatic; antiarthritic; vulnery; analgesic;	
KW	antiinflammatory; antibacterial; immunosuppressive; antiparkinsonian;	
KW	neuroprotective; nootropic; osteopathic; haemostatic; vasotropic;	
KW	antitumor; fungicide; antidiabetic; antiasthmatic; antiallergic;	
KW	immunostimulant; antiparasitic; secreted protein; transmembrane protein;	
KW	cytokine; cell proliferation; cell differentiation; autoimmune disease;	
KW	stem cell; growth factor; nervous system disease; neuropathy;	
KW	Alzheimer's disease; Parkinson's disease; Huntington's disease;	
KW	osteoporosis; severe combined immunodeficiency; SCID; infection;	
KW	multiple sclerosis; rheumatoid arthritis; gene therapy; gene; ss.	
XX		
OS	Homo sapiens.	
XX		
PN	US2002065394-A1.	
XX		
PD	30-MAY-2002.	
XX		
PF	22-DEC-2000; 2000US-0745763.	
XX		
PR	18-MAR-1998; 98US-0040963.	
XX		
PA	(JACO/) JACOBS K.	
PA	(MCCO/) MCCOY J M.	
PA	(LAVA/) LAVALLE E R.	
PA	(COLL/) COLLINS-RACIE L A.	



XX 28-JUL-2000; 2000EP-0116126.  
XX  
XX 29-JUL-1999; 99JP-0248036.  
XX 27-AUG-1999; 99JP-0300253.  
PR 11-JUN-2000; 2000JP-0118776.  
PR 02-MAY-2000; 2000JP-0183767.  
PR 09-JUN-2000; 2000JP-0241899.  
XX  
XX (HELI-) HELIX RES INST.  
PI Oca T, Isogai T, Nishikawa T, Hayashi K, Saito K, Yamamoto J,  
PI Ishii S, Sugiyama T, Wakamatsu A, Nagai K, Otsuki T;  
XX WPI; 2001-318749/34.  
XX  
XX Primer sets for synthesizing polynucleotides, particularly the 5602  
PT full-length cDNAs defined in the specification, and for the detection  
PT and/or diagnosis of the abnormality of the proteins encoded by the  
PT full-length cDNAs -  
XX  
XX Claim 8; SEQ ID 17952; 2537bp + CD ROM; English.  
XX  
XX The present invention describes primer sets for synthesizing 5602  
CC full-length cDNAs defined in the specification. Where a primer set  
CC comprises: (a) an oligo-dt primer and an oligonucleotide complementary  
CC to the complementary strand of a polynucleotide which comprises one of  
CC the 5602 nucleotide sequences defined in the specification, where the  
CC oligonucleotide comprises at least 15 nucleotides; or (b) a combination  
CC of an oligonucleotide comprising a sequence complementary to the  
CC complementary strand of a polynucleotide which comprises a 5'-end  
CC sequence and an oligonucleotide comprising a sequence complementary to a  
CC polynucleotide which comprises a 3'-end sequence, where the  
CC oligonucleotide comprises at least 15 nucleotides and the combination of  
CC the 5'-end sequence/3'-end sequence is selected from those defined in  
CC the specification. The primer sets can be used in antisense therapy and  
CC in gene therapy. The primers are useful for synthesizing polynucleotides,  
CC particularly full-length cDNAs. The primers are also useful for the  
CC detection and/or diagnosis of the abnormality of the proteins encoded by  
CC the full-length cDNAs. The primers allow obtaining of the full-length  
CC cDNAs easily without any specialised methods. AAH03166 to AAH13628 and  
CC AAH18633 to AAH18742 represent human cDNA sequences; AAH92446 to  
CC AAH95893 represent human amino acid sequences; and AAH13629 to AAH13632  
CC represent oligonucleotides, all of which are used in the exemplification  
CC of the present invention.  
XX  
XX Sequence 2178 BP; 525 A; 513 C; 678 G; 462 T; 0 other;  
SQ  
Query Match 99.6%; Score 1075.2; DB 22; Length 2178;  
Best Local Similarity 99.7%; Pred. No. 1.3e-212;  
Matches 1077; Conservative 0; Mismatches 3; Indels 0; Gaps 0;  
QY 1 ATGGCCGAGCATTCTTGTGAGAAATATCAACACAGCTCAAACTAGCAACTGTACAGT 60  
DB 716 ATGGCCGAGCATTCTTGTGAGAAATATCAACACAGCTCAAACTAGCAACTGTACAGT 775  
QY 61 GCTGCTGCTTCCAGAGAGCTGAACCTGAGCGCCCCCAGCGCGGAGAGCGGCTG 120  
DB 776 GCTGCTGCTTCCAGAGAGCTGAACCTGAGCGCCCCCAGCGCGGAGAGCGGCTG 835  
QY 121 CCCGAG 180  
DB 836 CCCGAG 895  
QY 181 CCCGAG 240  
DB 896 CCCGAG 955  
QY 241 GAATCTAGCTGCTGAG 300  
DB 956 GAATCTAGCTGCTGAG 1015  
QY 301 GGGGACTTCCCGCGCGGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 360

DB 1016 GGGGACTTCCCGCGCGGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1075  
QY 361 CTTTGTATGACTCCGAG 420  
DB 1076 CTTTGTATGACTCCGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1135  
QY 421 TGGGAGACAGCAG 480  
DB 1136 TGGGAGACAGCAG 1195  
QY 481 AAGCGGCTTGGAAACCGTACTCAAGCTGAACTGGGAAAGAGAGAGAGAGAGAGAG 540  
DB 1196 AAGCGGCTTGGAAACCGTACTCAAGCTGAACTGGGAAAGAGAGAGAGAGAGAG 1255  
QY 541 AAGCGGCTTGGAAACCGTACTCAAGCTGAACTGGGAAAGAGAGAGAGAGAGAGAG 600  
DB 1256 AAGCGGCTTGGAAACCGTACTCAAGCTGAACTGGGAAAGAGAGAGAGAGAGAG 1315  
QY 601 GGGCCCTATTAACAGCAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 660  
DB 1316 GGGCCCTATTAACAGCAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1375  
QY 661 AAAACCGGCTTGTACTTCCAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 720  
DB 1376 AAAACCGGCTTGTACTTCCAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1435  
QY 721 TTTCATAG 780  
DB 1436 TTTCATAG 1495  
QY 781 GAGTTTCTGACAG 840  
DB 1496 GAGTTTCTGACAG 1555  
QY 841 AACATGAG 900  
DB 1556 AACATGAG 1615  
QY 901 ATGGAG 960  
DB 1616 ATGGAG 1675  
QY 961 CGTGTGCGGAG 1020  
DB 1676 CGTGTGCGGAG 1735  
QY 1021 ACCGAG 1080  
DB 1736 ACCGAG 1795  
RESULT 5  
ABL90067/c  
ID ABL90067 standard; cDNA; 1083 BP.  
XX  
XX ABL90067;  
XX  
XX 24-MAY-2002 (first entry)  
XX  
XX Human polynucleotide SEQ ID NO 629.  
XX  
XX Cytostatic; immunosuppressive; nocotropic; neuroprotective; antiviral;  
XX antiallergic; hepatotropic; antidiabetic; antiinflammatory; antileuk;  
XX vulnary; anticonvulsant; antibacterial; antitumor; antiparasitic;  
XX cardiant; gene therapy; cancer; immune disorder; cardiovascular disorder;  
XX neurological disease; infection; human; secreted protein; gene; ss.  
OS Homo sapiens.  
XX  
XX PN WO200190304-A2.  
XX  
XX 29-NOV-2001.



XX 18-MAY-2001; 2001MO-US16450.  
 XX 19-MAY-2000; 2000US-20551SP.  
 XX (HUMA-) HUMAN GENOME SCI INC.  
 PA Birse CE, Rosen CA;  
 XX WPI; 2002-122018/16.  
 DR P-PSDB; ABB89658.  
 XX  
 PT Novel 1405 isolated polypeptides, useful for diagnosis, treatment and  
 PT prevention of neural, immune system, muscular, reproductive,  
 PT gastrointestinal, pulmonary, cardiovascular, renal and proliferative  
 PT disorders -  
 XX  
 PS Claim 4; SEQ ID NO 629; 2081bp + Sequence Listing; English.  
 XX  
 CC The invention relates to novel genes (ABL99449-ABL90853) and proteins  
 CC (ABB89040-ABB90444) useful for preventing, treating or ameliorating  
 CC medical conditions e.g. by protein or gene therapy. The genes are  
 CC isolated from a range of human tissues disclosed in the specification.  
 CC The nucleic acids, proteins, antibodies and (ant)agonists are useful  
 CC in the diagnosis, treatment and prevention of: (a) cancer, e.g. breast  
 CC and ovarian cancer and other cancers of the adrenal gland, bone, bone  
 CC marrow, breast, gastrointestinal tract, liver, lung, or urogenital;  
 CC (b) immune disorders e.g. Addison's disease, allergies, autoimmune  
 CC hemolytic anaemia, autoimmune thyroiditis, diabetes mellitus, Crohn's  
 CC disease, multiple sclerosis, rheumatoid arthritis and ulcerative  
 CC colitis; (c) cardiovascular disorders such as myocardial ischemias;  
 CC (d) wound healing; (e) neurological diseases e.g. cerebral anoxia and  
 CC epilepsy; and (f) infectious diseases such as viral, bacterial, fungal  
 CC and parasitic infections.  
 CC Note: The sequence data for this patent did not form part of the  
 CC printed specification, but was obtained in electronic format directly  
 CC from WIPO at ftp.wipo.int/pub/published\_pct\_sequences.  
 CC  
 XX Sequence 1083 BP; 301 A; 290 C; 199 G; 286 T; 7 other;  
 SQ  
 Query Match 40.2%; Score 434.2; DB 24; Length 1083;  
 Best Local Similarity 98.9%; Pred. No. 2.2e-80;  
 Matches 444; Conservative 3; Mismatches 1; Indels 1; Gaps 1;  
 QY 632 ATGATCAGACGACGAGAGCCGATCTCAAAACCGGCTGTACTCCAGCGGCGCG 691  
 DB 1083 ATGATCAGACGACGAGAGCCGATCTCAAAACCGGCTGTACTCCAGCGGCGCG 1024  
 QY 692 CCAATTCGACGACGACGATGACGATTCATGAAAGGGGGTGAAGAGATGGGG 751  
 DB 1023 -CAATTCGACGACGACGATGACGATTCATGAAAGGGGGTGAAGAGATGGGG 965  
 QY 752 GCGAGCATGGGATGGGAGGGGAGCGAGGTTTCGACGGGACTTCTCGAGAGCGT 811  
 DB 964 GCGAGCATGGGATGGGAGGGGAGCGAGGTTTCGACGGGACTTCTCGAGAGCGT 905  
 QY 812 ACGAGCGGTACACACGAGAGGCTTCGAGAAATGAGCAAGAGGCTCATCAAGAGT 871  
 DB 904 ACGAGCGGTACACACGAGAGGCTTCGAGAAATGAGCAAGAGGCTCATCAAGAGT 845  
 QY 872 ACCTGGAATGAGAGAGTCTCTCGCGCATGAGAGCGAACAACCGGCTCGGCTGG 931  
 DB 844 ACCTGGAATGAGAGAGTCTCTCGCGCATGAGAGCGAACAACCGGCTCGGCTGG 785  
 QY 932 AGAGCAAGCGGCTGGGAGGAGGAGCGCGGCTGGCGGAGCTGGAGCTGGAGC 991  
 DB 784 AGAGCAAGCGGCTGGGAGGAGGAGCGCGGCTGGCGGAGCTGGAGCTGGAGC 725  
 QY 992 GGTGCGCGCGGAGAACTCTCAGCTGTGACCGAGAAAGAACTGCAACCGGACAGAGG 1051  
 DB 724 GGTGCGCGCGGAGAACTCTCAGCTGTGACCGAGAAAGAACTGCAACCGGACAGAGG 665  
 QY 1052 GAGCGCGGCTTCCAAAGTTTGAGACTAG 1080

DB 664 GAGCGCGGCTTCCAAAGTTTGAGACTAG 636  
 RESULT 6  
 AAV86036  
 ID AAV86036 standard; cDNA; 523 BP.  
 XX  
 AC AAV86036;  
 XX  
 DT 27-APR-1999 (first entry)  
 XX  
 DE EST clone B227.  
 XX  
 XX Expressed sequence tag; secreted protein; haematopoiesis regulator;  
 XX tissue growth; activin; inhibin; tumour invasion suppressor; EST; human;  
 XX chemotaxis; chemokinesis; haemostasis; gene therapy; thrombolysis;  
 XX receptor; ligand; anti-inflammatory; tumour inhibitor; ds.  
 XX  
 OS Homo sapiens.  
 OS  
 PN MO9845435-A2.  
 XX  
 PD 15-OCT-1998.  
 XX  
 PF 10-APR-1998; 98MO-US06954.  
 XX  
 PR 10-APR-1997; 97US-0835913.  
 XX  
 PA (GENM) GENETICS INST INC.  
 XX  
 PI Agostino MJ, Jacobs K, Lavallie ER, McCoy JM, Merberg D;  
 PI Racle LA, Spalding V, Treacy M;  
 XX  
 DR WPI; 1999-070076/06.  
 XX  
 PT New polynucleotides encoding human secreted proteins - derived from  
 PT e.g. human blood, kidney, foetal lung, placenta, testes, brain,  
 PT ovary, pituitary, retina and colon cDNA libraries  
 XX  
 PS Claim 1; Page 97; 633bp; English.  
 XX  
 CC This sequence represents an expressed sequence tag (EST), and is a  
 CC polynucleotide of the invention. The polynucleotides of the invention are  
 CC all secreted EST sequences isolated from a variety of human tissue  
 CC sources. The EST sequences and proteins encoded by them are predicted to  
 CC have useful biological activities which would make them suitable for  
 CC treating, preventing or ameliorating medical conditions in humans and  
 CC animals, although no supporting data is given. Suggested activities  
 CC include nutritional activity, immune stimulating or suppressing activity,  
 CC haematopoiesis regulating activity, tissue growth activity,  
 CC activin/inhibin activity, chemotactic/chemokinetic activity, haemostatic  
 CC and thrombolytic activity, receptor/ligand activity, anti-inflammatory  
 CC activity, cadherin/tumour invasion suppressor activity, tumour inhibition  
 CC activity. The EST sequences are also stated to be useful for gene  
 CC therapy.  
 CC  
 XX Sequence 523 BP; 112 A; 149 C; 153 G; 108 T; 1 other;  
 SQ  
 Query Match 36.6%; Score 395.8; DB 20; Length 523;  
 Best Local Similarity 96.6%; Pred. No. 1.6e-72;  
 Matches 403; Conservative 1; Mismatches 13; Indels 0; Gaps 0;  
 QY 1 ATGGCCGAGCCATTCTTGCAGAAATCAACACGAGCTCAAACTGCACTGACAGT 60  
 DB 104 ATGGCCGAGCCATTCTTGCAGAAATCAACACGAGCTCAAACTGCACTGACAGT 163  
 QY 61 GCTGCTGCTGTCAGAGAGAGTGAACCTGAGCGGCCCCGAGCGCGGAGAGCGGCTG 120  
 DB 164 GCTGCTGCTGTCAGAGAGAGTGAACCTGAGCGGCCCCGAGCGCGGAGAGCGGCTG 223  
 QY 121 CCGAGAGAGAGAGTGGTGGCAATGAGAGGCTTCCCAAGTTGGGTGGCGCTCCGGGG 180



Db 224 CCCGAGAGACAGTAGTGCGCAATGAGAGCGTTCCTCCCAAGTTGGTGCGCTCCGGG 283  
 Qy 181 CCGAGAGGGGAGAGAGAGCTGTGATCCCAACCACTCTCCCTTGAGAGACCGAGGCTGTCA 240  
 Db 284 CCGAGAGGGGAGAGAGAGCTGTGATCCCAACCACTCTCCCTTGAGAGACCGAGGCTGTCA 343  
 Qy 241 GAATCTAGCTGCTTGAGAGAGGGCGAGAAAGGGCCAGAAATGGGGACGACTGTCCGTGCG 300  
 Db 344 GAATCTAGCTGCTTGAGAGAGGGCGAGAAAGGGCCAGAAATGGGGACGACTGTCCGTGCG 403  
 Qy 301 GGGGATTTCCCGCCCGCCGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 360  
 Db 404 GGGGATTTCCCGCCCGCCGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 463  
 Qy 361 CCTTGTGATGACTCCGAGGCGAGTAAAGTTGGGGGCTCTGCGCGAGGGGCGAGAGAG 417  
 Db 464 CCTTGTGATGACTCCGAGGCGAGTAAAGTTGGGGGCTCTGCGCGAGGGGCGAGAGAG 520  
 RESULT 7  
 ID ABX54966 standard; cDNA; 414 BP.  
 AC ABX54966;  
 XX  
 DT 26-FEB-2003 (first entry)  
 XX  
 DE Bovine EST associated with lactation/muscle/fat deposition #4895.  
 KW Bovine; ss; EST; expressed sequence tag; lactation; LMFD;  
 KM muscle deposition; fat deposition; genome mapping; gene identification;  
 XX gene analysis; cattle breeding.  
 OS Bos Taurus.  
 XX  
 PN US2002137160-A1.  
 XX  
 PD 26-SEP-2002.  
 XX  
 PF 26-OCT-2001; 2001US-0983965.  
 PR 17-DEC-1998; 98US-113678P.  
 PR 15-DEC-1999; 99US-0465231.  
 XX  
 PA (BYAT/) BYATT J C.  
 PA (MATH/) MATHIALAGAN N.  
 PA (TAON/) TAO N.  
 PA (WARR/) WARREN W C.  
 PI Byatt JC, Mathialagan N, Tao N, Warren WC;  
 DR WPI, 2003-102386/09.  
 XX  
 PT Purified nucleic acid molecules, useful for genome mapping, gene  
 PT identification and analysis, cattle breeding or preparation of  
 PT constructs for cattle gene expression and genetically improved cattle -  
 XX  
 PS Claim 2; SEQ ID No 4895; 389p; English.  
 XX  
 CC The invention relates to a purified nucleic acid molecule associated with  
 CC lactation or muscle and fat deposition (designated LMFD), derived  
 CC from cattle, and the LMFD nucleic acid can specifically hybridise to a  
 CC second nucleic acid molecule comprising any of 5912 nucleotide  
 CC sequences, appearing as ABX50072-ABX53983, or complements of them.  
 CC Also included are; (1) a transformed cell having a nucleic acid  
 CC comprising an LMFD nucleic acid linked to a promoter and a 3' non-  
 CC translated sequence that functions in the cell to cause termination of  
 CC transcription and addition of polyadenylated ribonucleotides to a 3' end  
 CC of the mRNA molecule; and (2) determining a level or pattern of a  
 CC molecule in a bovine cell or tissue comprising: (a) incubating a marker  
 CC nucleic acid (comprising any of the 5912 nucleic acid sequences or its  
 CC complement or fragment) with a complementary nucleic acid molecule  
 CC obtained from the bovine cell or tissue, where hybridisation between the

CC marker nucleic acid and the complementary nucleic acid permits the  
 CC detection of the molecule; and (b) detecting the level or pattern of the  
 CC complementary nucleic acid, where the detection of the complementary  
 CC nucleic acid is predictive of the level or pattern of the molecule.  
 CC The LMFD nucleic acid is used for determining a level or pattern  
 CC of a molecule in a bovine cell or tissue. It is useful for genome  
 CC mapping, gene identification and analysis, cattle breeding, preparation  
 CC of constructs for use in cattle gene expression, or for genetically  
 CC improving cattle. The present sequence is one of the 5912 bovine  
 CC LMFD EST (expressed sequence tag) nucleic acids.  
 CC Note: The present sequence was not shown in the specification but  
 CC was obtained in electronic format from the USPTO web site:  
 CC [seqdata.uspto.gov/Sequence.html?DocID=20020137160](http://seqdata.uspto.gov/Sequence.html?DocID=20020137160).  
 XX  
 SQ Sequence 414 BP; 117 A; 106 C; 130 G; 61 T; 0 other;  
 Query Match 33.6%; Score 362.8; DB 25; Length 414;  
 Best Local Similarity 92.3%; Pred. No. 9.9e-66;  
 Matches 382; Conservative 0; Mismatches 32; Indels 0; Gaps 0;  
 Qy 462 GAGAGCGCCGTCAGAGAAAGCGGCAATTGAAACCGTACTACAGCTGAACTGGGAGA 521  
 Db 1 GAGAGCGCCGTCAGAGAAAGCGGCAATTGAAACCGTACTACAGCTGAACTGGGAGA 60  
 Qy 522 GAAGAAAAAGTTGAGAGAGAAAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 581  
 Db 61 GAAGAAAAAGTTGAGAGAGAAAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 120  
 Qy 582 CGCCAAAGGGCCAGCGCGGTCGCGCCTTATACACACAGAGTTCTCTATGATGATCAGCA 641  
 Db 121 CGCCAAAGGGCCAGCGCGGTCGCGCCTTATACACACAGAGTTCTCTATGATGATCAGCA 180  
 Qy 642 CCAGAGAGAGAGCGGATCTCAAAACCGGCTGTACTCCAAAGCGGCGCGCCAAATCCGA 701  
 Db 181 CCAGAGAGAGAGCGGATCTTAAACCGGCTGTACTCCAAAGCGGCGCGCCAAATCCGA 240  
 Qy 702 CGACACAGCGATGAG 761  
 Db 241 CGACACAGCGATGAG 300  
 Qy 762 GATGGAGAGGAGAGCGGAG 821  
 Db 301 GATGGAGAGGAGAGCGGAG 360  
 Qy 822 CCACACGAGAGCGCTGCAAGATGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 875  
 Db 361 CCACACGAGAGCGCTGCAAGATGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 414  
 RESULT 8  
 ID ABL63543 standard; DNA; 461 BP.  
 AC ABL63543;  
 XX  
 DT 15-MAY-2002 (first entry)  
 XX  
 DE Breast cancer related gene sequence SEQ ID NO:1880.  
 XX  
 KW Human; cancer; colon; breast; ovary; oesophagus; kidney; thyroid;  
 KM stomach; lung; prostate; pancreas; carcinoma; antitumour; cancerous;  
 KW cytostatic; gene therapy; antineoplastic; Wilms' tumour; adenocarcinoma;  
 XX gene; ds.  
 OS Homo sapiens.  
 XX  
 PN WO200194629-A2.  
 XX  
 PD 13-DEC-2001.  
 XX  
 PF 30-MAY-2001; 2001WO-US10838.  
 XX  
 PR 05-JUN-2000; 2000US-209473P.

PR 05-JUN-2000; 2000US-209531P.  
 PR 18-SEP-2000; 2000US-233133P.  
 PR 18-SEP-2000; 2000US-233617P.  
 PR 20-SEP-2000; 2000US-234009P.  
 PR 20-SEP-2000; 2000US-234034P.  
 PR 20-SEP-2000; 2000US-234052P.  
 PR 22-SEP-2000; 2000US-234505P.  
 PR 22-SEP-2000; 2000US-234567P.  
 PR 25-SEP-2000; 2000US-234923P.  
 PR 25-SEP-2000; 2000US-234924P.  
 PR 25-SEP-2000; 2000US-235077P.  
 PR 25-SEP-2000; 2000US-235082P.  
 PR 25-SEP-2000; 2000US-235134P.  
 PR 25-SEP-2000; 2000US-235280P.  
 PR 26-SEP-2000; 2000US-235637P.  
 PR 26-SEP-2000; 2000US-235638P.  
 PR 27-SEP-2000; 2000US-235711P.  
 PR 27-SEP-2000; 2000US-235720P.  
 PR 27-SEP-2000; 2000US-235840P.  
 PR 27-SEP-2000; 2000US-235863P.  
 PR 28-SEP-2000; 2000US-236028P.  
 PR 28-SEP-2000; 2000US-236032P.  
 PR 28-SEP-2000; 2000US-236033P.  
 PR 28-SEP-2000; 2000US-236034P.  
 PR 28-SEP-2000; 2000US-236109P.  
 PR 29-SEP-2000; 2000US-236111P.  
 PR 29-SEP-2000; 2000US-236842P.  
 PR 29-SEP-2000; 2000US-236891P.  
 PR 02-OCT-2000; 2000US-237172P.  
 PR 02-OCT-2000; 2000US-237173P.  
 PR 02-OCT-2000; 2000US-237278P.  
 PR 02-OCT-2000; 2000US-237294P.  
 PR 02-OCT-2000; 2000US-237295P.  
 PR 02-OCT-2000; 2000US-237316P.  
 PR 03-OCT-2000; 2000US-237425P.  
 PR 03-OCT-2000; 2000US-237598P.  
 PR 03-OCT-2000; 2000US-237604P.  
 PR 03-OCT-2000; 2000US-237606P.  
 PR 03-OCT-2000; 2000US-237608P.  
 PR 01-NOV-2000; 2000US-244867P.  
 PR 01-NOV-2000; 2000US-245084P.  
 XX  
 PA (AVAL-) AVALON PHARM.  
 XX  
 PI Young PE, Augustus M, Carter KC, Ebner R, Endress G, Horrigan S,  
 PI Soppet DR, Weaver Z;  
 XX  
 XX WPI; 2002-188264/24.  
 DR  
 XX  
 XX  
 PT Screening for anti-neoplastic agent involves exposing cells to a  
 PT chemical agent to be tested for anti-neoplastic activity, and  
 PT determining a change in expression of a gene of a signature gene set -  
 XX  
 OS  
 PS Claim 1; SEQ ID 1880; 44bp; English.  
 XX  
 XX  
 CC The present invention describes a method (M1) for screening for an  
 CC anti-neoplastic agent. The method involves exposing cells to a chemical  
 CC agent to be tested for anti-neoplastic activity, determining a change in  
 CC expression of at least one gene (I) of a signature gene set, where (I)  
 CC comprises a sequence (S) selected from 8447 sequences (given in ABL61664  
 CC to ABL70110), or is at least 95% identical to (S), where a change in  
 CC expression is indicative of anti-neoplastic activity. (I) has cytostatic  
 CC activity and can be used in gene therapy. M1 can be used for screening  
 CC an anti-neoplastic agent, and can be used for producing a product which  
 CC is the data collected with respect to the anti-neoplastic agent as a  
 CC result of M1, and the data is sufficient to convey the chemical  
 CC structure and/or properties of the agent. M1 can be used in the  
 CC treatment of cancer such as colon, breast, stomach, lung, thyroid,  
 CC oesophageal, ovarian, kidney, prostate or pancreatic cancer,  
 CC adenocarcinoma, carcinoma, clear cell cancer, infiltrating ductal cancer,  
 CC infiltrating lobular cancer, squamous cell carcinoma, neuroendocrine  
 CC carcinoma, papillary carcinoma and Wilms' tumour.  
 CC  
 XX

Sequence 461 BP; 105 A; 134 C; 138 G; 80 T; 4 other;  
 Query Match 32.9%; Score 355.2; DB 24; Length 461;  
 Best Local Similarity 95.7%; Pred. No. 3,7e-64;  
 Matches 374; Conservative 0; Mismatches 16; Indels 1; Gaps 1;  
 QY 1 ATGGCCGAGCCATTCTTGTCAATATCAACCCAGCTCAAACTNAGCACTGTACAGGT 60  
 DB 72 ATGGCCGAGCCATTCTTGTCAATATCAACCCAGCTCAAACTNAGCACTGTACAGGT 131  
 QY 61 GCTGCTGCTGCTCCAGGAAGCTGAACCTGAGCCGCCGCCGAGGCGGAGGAGGCGGTG 120  
 DB 132 GCTGCTGCTGCTCCAGGAAGCTGAACCTGAGCCGCCGCCGAGGCGGAGGCGGTG 191  
 QY 121 CCCGAGGAGGAGAGTAGTAGTGAGATCGAGAGCGTTCCCGAGTTGGTGGCCGTCCGAGG 180  
 DB 192 CCCGAGGAGGAGAGTAGTAGTGAGATCGAGAGCGTTCCCGAGTTGGTGGCCGTCCGAGG 251  
 QY 181 CCGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGG 240  
 DB 252 CCGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGG 311  
 QY 241 GAATCTAGCTGCTGAGAGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGG 300  
 DB 312 GAATCTAGCTGCTGAGAGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGG 371  
 QY 301 GCGGACTTCCCGCCGCGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGG 360  
 DB 372 -CGGACTTCCCGCCGCGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGG 430  
 QY 361 CCTGTGATGACTCCGAGGCGCAGTAAGTTGG 391  
 DB 431 CCTGTGATGACTCCGAGGCGCAGTAAGTTGG 461  
 RESULT 9  
 ABQ61099  
 ID ABQ61099 standard; cDNA; 1557 BP.  
 XX  
 AC ABQ61099;  
 XX  
 XX 26-FEB-2003 (first entry)  
 DT  
 XX  
 XX Mouse putative protein #15 encoding sequence.  
 DE  
 XX  
 XX Neuroprotective; immunomodulator; cancer;  
 KW cytosolic; anti-inflammation; gene therapy; nutritional supplement;  
 KW wound; burn; ulcer; Alzheimer's disease; Huntington's disease;  
 KW amyotrophic lateral sclerosis; autoimmune disorder; inflammation;  
 KW vulnery; gene; ss.  
 XX  
 OS Mus musculus.  
 XX  
 PN WO200231111-A2.  
 XX  
 PD 18-APR-2002.  
 XX  
 PF 11-OCT-2001; 2001WO-US27760.  
 XX  
 PR 12-OCT-2000; 2000US-0687527.  
 XX  
 PA (HYSE-) HYSEQ INC.  
 PI Tang YT, Liu C, Zhou P, Asundi V, Zhang J, Zhao QA, Ren F;  
 PI Xue AJ, Yang Y, Wehrman T, Drmanac RT;  
 XX  
 DR WPI; 2002-426278/45.  
 XX  
 DR N-PSDB; ABP43855.  
 XX  
 PT New polypeptides and their encoded proteins, useful as nutritional  
 PT sources or supplements, or in gene therapy, particularly for treating  
 PT wounds, Alzheimer's disease, amyotrophic lateral sclerosis, cancer or  
 PT inflammation -  
 XX

XX Claim 1, SEQ ID # 312; 357pp + sequence listing; English.  
 PS The invention relates to 446 newly isolated polynucleotide sequences.  
 CC The activity of polynucleotides of the invention may be described as:  
 CC anti-inflammatory, immunomodulatory, cytoprotective and  
 CC anti-inflammatory. Compositions comprising nucleic acids of the invention  
 CC are useful for treating a mammalian subject, or as nutritional sources or  
 CC supplements. These are useful in gene therapy, particularly for treating  
 CC wounds, burns or ulcers, Alzheimer's disease, Huntington's disease,  
 CC amyotrophic lateral sclerosis, autoimmune disorders, cancer or  
 CC inflammation. The nucleic acids and polypeptides are also useful in  
 CC diagnostic and research methods. The sequences given in records  
 CC AB060788-AB061233 represent polynucleotides of the invention.  
 CC NOTE: The sequence data for this patent did not form part of the printed  
 CC specification, but was obtained in electronic format directly from WIPO  
 CC at ftp.wipo.int/pub/published\_pct\_sequences.

SQ Sequence 1557 BP; 355 A; 426 C; 513 G; 263 T; 0 other;

Query Match 16.2%; Score 174.8; DB 24; Length 1557;

Best Local Similarity 60.5%; Pred. No. 7.8e-27; Mismatches 197; Indels 39; Gaps 3;

Matches 362; Conservative 0; Mismatches 197; Indels 39; Gaps 3;  
 441 GCTGGGGAAGAAAAATATAGAGAGCCCGCTCCAGAGAGAGCGGCTTGGAAACCGTA 500  
 415 GCTGGGCGGAGAAACCCGTCGGCGGCATGAGAGCCCAAGGCACTGGGAGACCTTA 474  
 501 CTACAACTGAACTGGGAAAGAGAAAAAGTTGACGAGAAACAGAGCCCTTCAGCTTC 560  
 475 CTTGGAGCTGAGCTGGCTGAGAAACAGAGCGGAGTAGAGAGCAGAGCGAGGCGCTC 534  
 561 AAGATCCGAGCGGAGATGTTGCGCAAGGCGCGGCTGCGGCTTATTAACACAGCA 620  
 535 CCGGGTCGCGGAGAGATGTTGCGCAAGGCGCGGCTGCGGCTTATTAACACAGCA 594  
 621 GTTCTCATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 680  
 595 GTTCTCATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 646  
 681 GCGGGCGCGCGGCAATCCGAGCAGACGAGATGACGATGATGATGATGATGATGATGAT 740  
 647 GGGATCTCCACCCAGGTTTCAG-----TGGGAGAG 678  
 741 GGAGATGGGGGCGAGATGAGATGAGATGAGATGAGATGAGATGAGATGAGATGAGAT 800  
 679 TGAGGCGCGGAGACAGATGAGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCTT 738  
 801 CTGAGAGAGTACGAGCGGATGACACAGAGAGCTGACAGATGAGAGAGAGAGCT 860  
 739 CTCTGAGACTTACGAGCGCTTCCACACGAGAGCTGACAGAGCGGCGGCGGCGGAGCT 798  
 861 CATCAAGAGATGATGAGAACTGAGAAAGTCTCTGCGCATGAGAGAGAGAGCAACCG 920  
 799 GGTGCGAGACTACTGAGAGCTGAGAGAGCGGCTGTGCGAGCGAGAGAGAGAGAGCT 858  
 921 GCTGCGGCTGAGCT 977  
 859 GGTGCGAGAGCTGAGCT 918  
 978 GCTGAGAGTGAACGGGCTGCGCGCGAGAACTGACAGTGTGACGAGAGAGAGAGAGCT 1035  
 919 TGGCGAGTCCAGAGAGCTCCGAGACGAGAAACAGAGCGGCTTCTCAGAGAGAGAGAG 976

RESULT 10  
 ID AAV87316 standard; cDNA; 342 BP.

XX AAV87316;  
 AC  
 XX  
 DT 27-APR-1999 (first entry)  
 XX

DE EST clone BP783.

XX Expressed sequence tag; secreted protein; haematopoiesis regulator;  
 KW tissue growth; activin; inhibin; tumour invasion suppressor; EST; human;  
 KW chemotaxis; chemokinesis; haemostasis; gene therapy; thrombolytic;  
 KW receptor; ligand; anti-inflammatory; tumour inhibitor; ds.

OS Homo sapiens.

PN MO9845435-A2.

XX 15-OCT-1998.

PD 10-APR-1998; 98WO-US06954.

XX 10-APR-1997; 97US-0835913.

XX (GENV) GENETICS INST INC.

XX Agostino MJ, Jacobs K, Lavallie ER, McCoy JM, Merberg D;  
 PI Racine LA, Spaulding V, Treacy M;  
 XX WPI, 1999-070076/06.

XX New polynucleotides encoding human secreted proteins - derived from  
 PT e.g. human blood, kidney, foetal lung, placenta, testes, brain,  
 PT ovary, pituitary, retina and colon cDNA libraries

XX Claim 1; Page 526; 633pp; English.

XX This sequence represents an expressed sequence tag (EST), and is a  
 CC polynucleotide of the invention. The polynucleotides of the invention are  
 CC all secreted EST sequences isolated from a variety of human tissue  
 CC sources. The EST sequences and proteins encoded by them are predicted to  
 CC have useful biological activities which would make them suitable for  
 CC treating, preventing or ameliorating medical conditions in humans and  
 CC animals, although no supporting data is given. Suggested activities  
 CC include nutritional activity, immune stimulating or suppressing activity,  
 CC haematopoiesis regulating activity, tissue growth activity,  
 CC activin/inhibin activity, chemotactic/chemokinetic activity, haemostatic  
 CC and thrombolytic activity, receptor/ligand activity, anti-inflammatory  
 CC activity, cadherin/tumour invasion suppressor activity, tumour inhibition  
 CC activity. The EST sequences are also stated to be useful for gene  
 CC therapy.

XX Sequence 342 BP; 90 A; 79 C; 83 G; 89 T; 1 other;

Query Match 12.8%; Score 138; DB 20; Length 342;

Best Local Similarity 100.0%; Pred. No. 2.4e-19; Mismatches 0; Indels 0; Gaps 0;

Matches 138; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
 1 ATGGCGGAGCCATTCTTTCAGAAATATCAACACGAGCTCAAACTGAACTGACAGT 60  
 196 ATGGCGGAGCCATTCTTTCAGAAATATCAACACGAGCTCAAACTGAACTGACAGT 255  
 61 GCTGCTGCTGTCAGAGAGAGTGAACCTGAGCGGCGGCGGCGGAGAGAGCGGCTG 120  
 256 GCTGCTGCTGTCAGAGAGAGTGAACCTGAGCGGCGGCGGCGGAGAGAGCGGCTG 315  
 121 CCCGAGAGAGAGAGTACG 138  
 316 CCCGAGAGAGAGAGTACG 333

RESULT 11  
 ID AAH9667/c standard; cDNA; 997 BP.

XX AAH9667;  
 AC  
 XX  
 DT 16-OCT-2001 (first entry)  
 XX Human protein encoding cDNA sequence SEQ ID NO:502.  
 DE

XX Human; cancer; ulcer; HIV infection; human immunodeficiency virus;  
 KW antiinflammatory; antirheumatic; antiarthritic; immunosuppressive;  
 KW antiinfective; endocrine; cardiant; central nervous system; virude;  
 KW anti-HIV; fungicide; antimutagen; cardiovascular; antineoplastic; anaemia;  
 KW antiparasitic; haemostatic; vulnery; antilucer; osteopathic; eczema;  
 KW dermatological; antiallergic; antistatic; antidiabetic; cytosolic;  
 KW neuroprotective; antidepressant; nootropic; antiparkinsonian; infection;  
 KW immunostimulant; gene therapy; antisense therapy; vaccine; inflammation;  
 KW immunohyaline; rheumatoid arthritis; septic shock; pancreatitis;  
 KW cardiac dysfunction; neuropathology; cardiac anaphylaxis; autoimmunity;  
 KW genetic disease; haematopoietic disorder; platelet disorder; asthma;  
 KW chromocytoma; osteoporosis; severe combined immunodeficiency;  
 KW allergic rhinitis; diabetes; multiple sclerosis; depression;  
 KW Alzheimer's disease; Parkinson's disease; neurodegenerative disorder;  
 KW neurological disorder; ss.

OS Homo sapiens.

PN WO200153455-A2.

PD 26-JUL-2001.

PF 22-DEC-2000; 2000MO-US35017.

PR 23-DEC-1999; 99US-0471275.

PR 21-JAN-2000; 2000US-048725.

PR 25-APR-2000; 2000US-0552317.

XX (HYSE-) HYSEO INC.

PI Tang YT, Liu C, Drmanac RT;

DR WPI; 2001-457603/49.

DR P-PSDB; AAM25726.

XX Isolated human polynucleotides encoding polypeptides, useful for the  
 PT treatment and diagnosis of e.g. cancer, ulcers and HIV infection -  
 XX Claim 1; Page 566-567; 1217pp; English.

XX AAH99166 to AAH99904 encode the human proteins given in AAM25225 to  
 CC AAM25963. The proteins can have activities based on the tissues and  
 CC cells they are expressed in, such as: antiinflammatory; antirheumatic;  
 CC antiallergic; immunosuppressive; antibacterial; endocrine; cardiant;  
 CC central nervous system; virude; anti-HIV; fungicide; antimutagen;  
 CC cardiovascular; antineoplastic; antiparasitic; haemostatic; vulnery;  
 CC antilucer; osteopathic; dermatological; antiallergic; antistatic;  
 CC antidiabetic; cytosolic; neuroprotective; antidepressant; nootropic;  
 CC antiparkinsonian; and immunostimulant. The proteins and polynucleotides  
 CC encoding them can be used in gene therapy, antisense therapy and vaccine  
 CC production. The proteins and polynucleotides are useful for screening for  
 CC agonists or antagonists of a protein and for the treatment and diagnosis  
 CC of disorders associated with the activity of a protein e.g. inflammation,  
 CC rheumatoid arthritis, septic shock, pancreatitis, cardiac dysfunction,  
 CC neuropathology, cardiac anaphylaxis, viral, bacterial, HIV and fungal  
 CC infections, autoimmunity, genetic diseases, haematopoietic disorders,  
 CC anaemia, platelet disorders, thrombocytopenia, wounds, burns, ulcers,  
 CC osteoporosis, severe combined immunodeficiency, eczema, allergic  
 CC rhinitis, asthma, diabetes, cancer, multiple sclerosis, depression,  
 CC Alzheimer's disease, Parkinson's disease, neurodegenerative and  
 CC neurological disorders.

XX Sequence 997 BP; 184 A; 323 C; 266 G; 224 T; 0 other;

Query Match 10.4%; Score 112.4; DB 22; Length 997;

Best Local Similarity 62.7%; Pred. No. 5.4e-14;

Matches 192; Conservative 0; Mismatches 111; Indels 3; Gaps 1;

DB 733 GGGGGGAGAGATGGGGGAGAGATGGGGGAGAGAGAGAGAGATTTGGAG 792  
 887 GGGGAGAGTGGGCGGAGACAGTGTGGGGGGGGGCGGAGCGACGAGTTCAGCG 828

QY 793 CCGGACTTCTCGAGAGCTAGCAGCGGCTACACACGAGAGCCTTGCAACATGAGCAG 852  
 DB 827 AAGGACTTCTCGAGAGCTTACAGAGCTTCCACACGAGAGCCTTGAGCGCGCAGCAAG 768  
 QY 853 CAGGAGCTCATGAGAGAGTACCTGGAAGTGGAGTCTCGGCGATGAGAGCAG 912  
 DB 767 CAGGAGCTCTGCGAGACTGAGCTGAGAGAGAGAGAGAGAGAGAGAGAGAG 708  
 QY 913 AACACCGGCTCGGCTGAGAGAGCAAGCGGCTGGTGCGA---CGACGCGCTGTGCGG 969  
 DB 707 ACTAGAGAGCTGAGAGAGCTGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 648  
 QY 970 GAGCTGAGAGCTGAGAGTGGACCGGCTGCGCGCGGAGAACTTCACTGTCAGCAGAAC 1029  
 DB 647 GAGCTGCGTCCGAGAGCTCAGAGAGCTCCGAGCCGAAACAGAGAGAGAGAGAG 588  
 QY 1030 GAAGCTG 1035  
 DB 587 CAGATG 582

RESULT 12

AAH12885/C

ID AAH12885 standard; cDNA; 567 BP.

AC AAH12885;

DT 26-JUN-2001 (first entry)

DE Human cDNA clone (3'-primer) SEQ ID NO:9720.

XX Human; primer; detection; diagnosis; antisense therapy; gene therapy; ss.

XX Homo sapiens.

PN EP1074617-A2.

PD 07-FEB-2001.

PF 28-JUL-2000; 2000EP-0116126.

PR 29-JUL-1999; 99JP-0248036.

PR 27-AUG-1999; 99JP-0300253.

PR 11-JAN-2000; 2000JP-0118776.

PR 02-MAY-2000; 2000JP-0183767.

PR 09-JUN-2000; 2000JP-0241899.

XX (HELI-) HELIX RES INST.

PI Ota T, Isogai T, Nishikawa T, Hayaashi K, Saito K, Yamamoto J;

PI Ishii S, Sugiyama T, Wakamatsu A, Nagai K, Otsuki T;

XX WPI; 2001-318749/34.

DR Primer sets for synthesizing polynucleotides, particularly the 5602

PT full-length cDNAs defined in the specification, and for the detection

PT and/or diagnosis of the abnormality of the proteins encoded by the

PT full-length cDNAs -

XX Claim 3; SEQ ID 9720; 2537bp + CD ROM; English.

XX The present invention describes primer sets for synthesizing 5602

CC full-length cDNAs defined in the specification. Where a primer set

CC comprises: (a) an oligo-dT primer and an oligonucleotide complementary

CC to the complementary strand of a polynucleotide which comprises one of

CC the 5602 nucleotide sequences defined in the specification, where the

CC oligonucleotide comprises at least 15 nucleotides; or (b) a combination

CC of an oligonucleotide comprising a sequence complementary to the

CC complementary strand of a polynucleotide which comprises a 5'-end

CC sequence and an oligonucleotide comprising a sequence complementary to a

CC polynucleotide which comprises a 3'-end sequence, where the

CC oligonucleotide comprises at least 15 nucleotides and the combination of

CC the 5'-end sequence/3'-end sequence is selected from those defined in

CC the specification. The primer sets can be used in antisense therapy and  
 CC in gene therapy. The primers are useful for synthesizing polynucleotides,  
 CC particularly full-length cDNAs. The primers are also useful for the  
 CC detection and/or diagnosis of the abnormality of the proteins encoded by  
 CC the full-length cDNAs. The primers allow obtaining of the full-length  
 CC cDNAs easily without any specialised methods. AAH03166 to AAH13628 and  
 CC AAH13633 to AAH18742 represent human cDNA sequences; AAB92446 to  
 CC AAB95893 represent human amino acid sequences; and AAH13629 to AAH13632  
 CC represent oligonucleotides, all of which are used in the exemplification  
 CC of the present invention.

XX Sequence 567 BP, 182 A; 139 C; 87 G; 149 T; 10 other;

Query Match 7.7%; Score 82.8; DB 22; Length 567;  
 Best Local Similarity 76.7%; Pred. No. 6.3e-08;  
 Matches 122; Conservative 0; Mismatches 34; Indels 3; Gaps 2;

QY 922 CTGCGGCTGAGAGCAACCGGCTGGTGCGACGACGCGGTGCGGAGCTGGAGCTG 981  
 Db 539 CNGTTTGGTTGAGACAAACGCTGGTGGGAAACCGACCTTTTGGCGGAG-TGGAGTTG 481

QY 982 GAGGTGACCGGCTGGCGCGCGGAGACCTGCACTGCTGACCGAGAACGACCTGCCG 1041  
 Db 480 GAGTTG-ACCGGTGCGCGCGCGGAGACCTTCCAGTTGAGACCGAGAACGACCTGCCG 423

QY 1042 CAGCAGAGGAGGAGCGCGCTTTCCAACTTTGAGAGACTAG 1080  
 Db 422 CAGCAGAGGAGGAGCGCGCGCTTTCCAACTTTGAGAGACTAG 384

RESULT 13  
 ID AAV55831 standard; DNA; 799 BP.

XX AAV55831;  
 DT 18-NOV-1998 (first entry)

XX Nucleotide sequence of the stabilising sequence-encoding insert.

XX Fusion protein; stabilising polypeptide; proteolytic degradation;  
 XX resistance; half-life; autoimmune disease; inflammation; nitro drug;  
 XX IkappaB regulator protein; inflammatory bowel disease; in vivo imaging;  
 XX nitroreductase protein; enzyme therapy; prodrug therapy; protease;  
 XX cancer; pathological condition; ss.

XX Epstein-barr virus.

XX WO9822577-A1.

XX 28-MAY-1998.

XX 17-NOV-1997; 97WO-IB01508.

XX 25-JUN-1997; 97US-0048945.

XX 15-NOV-1996; 96US-0030986.

XX (MASU/) MASUCCI M G.

XX Masucci MG;

XX WPI; 1998-312463/27.

XX New fusion proteins resistant to proteolytic degradation -  
 XX comprising a core protein with a stabilising polypeptide comprising  
 XX a peptidic sequence containing glycine repeats

XX Disclosure; Fig 4B; 120pp; English.

XX This is a nucleotide sequence of the stabilising sequence-encoding  
 XX insert. The invention provides a method for increasing the resistance  
 XX of a core protein to proteolytic degradation that comprises linking or  
 XX inserting onto or into the core protein a stabilising polypeptide of

CC formula (Gly)X(Glyb)Y(Glyc)Zln where Glya, Glyb, Glyc are 1-6  
 CC sequential Gly residues and X, Y, Z are Ala, Ser, Val, Ile, Leu, Met,  
 CC Phe, Pro or Thr and n can be anything between 1-66. X, Y and Z need not  
 CC be identical from n repeat to n repeat. Alternatively a nucleic acid  
 CC encoding the stabilising polypeptide can be linked onto or inserted into  
 CC a nucleic acid encoding a core protein. The fusion proteins of the  
 CC invention are more resistant to degradation by proteases and, thus, have  
 CC a longer half-life than the unfused core protein. The products can be  
 CC used for treating autoimmune diseases, cancer and inflammation. In  
 CC particular, the core protein may be an IkappaB regulator protein for the  
 CC treatment of inflammatory bowel disease, or a nitroreductase protein  
 CC which can activate nitro drugs in enzyme/prodrug therapy to treat cancer  
 CC or other pathological conditions. The fusion proteins can also be used in  
 CC diagnostic methods such as in vivo imaging.

XX Sequence 799 BP; 201 A; 106 C; 479 G; 13 T; 0 other;

Query Match 6.3%; Score 68.4; DB 19; Length 799;  
 Best Local Similarity 44.0%; Pred. No. 6.3e-05;  
 Matches 288; Conservative 0; Mismatches 366; Indels 0; Gaps 0;

QY 406 GGGGGGAGAGAGAGTGGGGGACAGACGACGACGCTGGGAGAAAACATTAAGAGA 465  
 Db 63 GGGCAGAGACGAGAGAGAGGGGCGAGAGCAGAGAGAGGGGCGAGAGGGGCGAG 122

QY 466 CGCCCTCCCAAGAGAGCGGCACTTGGAAACCGTACTCAACCTGMACTGGAGAGAG 525  
 Db 123 GAGGGGCAAGAGACAGAGAGAGGGGCGAGAGCAGAGAGAGGGGCGAGAGGGG 182

QY 526 AAAAGTTGACGAGAGAAACAGAGCTTCCAGCTTCAAGATCCAGACCGAGATGTTCCG 585  
 Db 183 CAGAGACAGAGAGAGGGGCGAGAGCAGAGAGGGGCGAGAGGGGCGAGAGAGAG 242

QY 586 AAGGGCCAGCCGGTGGCGGCTTATAACACGACGAGTTCTCATGTGATGATCAGCCAG 645  
 Db 243 GGGCAGAGAGGGGCGAGAGGGGCGAGAGCAGAGAGGGGCGAGAGAGAGAGGGGCG 302

QY 646 GAGAGGCGGATCTCAAAACCGGCTGTACTCAACGCGGCGCGCCCAATCCGACGAC 705  
 Db 303 GAGGGGCAAGAGACAGAGAGAGGGGCGAGAGGGGCGAGAGGGGCGAGAGAGAGGGG 362

QY 706 ACCAGGATGACGACTTCAATGAGAGAGGGGCTGAGAGAGATGGGGCAGCGATGGATG 765  
 Db 363 CAGAGACAGAGAGGGGCGAGAGGGGCGAGAGGGGCGAGAGAGAGAGAGAGAGAGAG 422

QY 766 GAGAGGAGAGCGAGAGAGTTTCTGACGCGGACCTTCTGGAACCTGACGGGTTCAC 825  
 Db 423 GAGGGGCGAGAGAGGGGCGAGAGGGGCGAGAGGGGCGAGAGAGAGAGAGAGAGAG 482

QY 826 ACGAGAGCCTGAGAAATGAGCAAGAGAGCTCATAGAGAGTCACTGAACTGAGAG 885  
 Db 483 CAGAGAGGGGCGAGAGAGAGGGGCGAGAGGGGCGAGAGAGAGAGAGAGAGAGAGAG 542

QY 886 AAGTGCTCTCGCGCATGAGAGAGAGAAACACCGGCTGCGCTGAGAGAGAGCGGCTG 945  
 Db 543 GAGCAGAGAGGGGCGAGAGGGGCGAGAGGGGCGAGAGGGGCGAGAGAGAGAGAGAG 602

QY 946 GGTGGCAGAGAGCGCGGTGTGCGGAGCTGAGAGCTGAGAGCTGAGAGCTGAGAGCT 1005  
 Db 603 GAGGGGCGAGAGAGGGGCGAGAGAGGGGCGAGAGGGGCGAGAGAGAGAGAGAGAGAG 662

QY 1006 AACCTCAGCTGTCGAGAGAGAGAAAGAACTGCAACCGGCGAGAGAGAGAGAGAG 1059  
 Db 663 GGGCAGAGAGAGAGAGGGGCGAGAGGGGCGAGAGGGGCGAGAGAGAGAGAGAGAGAG 716

RESULT 14  
 ID AAA50254 standard; DNA; 1926 BP.

XX AAA50254;  
 DT 07-NOV-2000 (first entry)

XX Epstein Barr virus nuclear antigen 1 protein (EBNA1) DNA.  
 DE EBV; nuclear antigen 1; EBNA1; episome; transfection; selection;  
 KM gene therapy; ds.  
 XX  
 OS Epstein-barr virus.  
 XX  
 PN MO20004778-A1.  
 PD 17-AUG-2000.  
 XX  
 PF 11-FEB-2000; 2000MO-US03547.  
 XX  
 PR 11-FEB-1999; 99US-0249585.  
 XX  
 PA (PHAR-) PHARMACOPEDIA INC.  
 PI Horlick RA, Chelsky D;  
 XX  
 DR WPI; 2000-515062/46.  
 DR P-PSDB; AAY95856.  
 XX  
 PT Stably transfecting eukaryotic cells with at least one episome for the  
 PT production of a desired protein in vitro and for gene therapy -  
 PS Disclosure; Fig 2; 53pp; English.

XX The present sequence is that of DNA encoding the Epstein-Barr virus  
 CC (EBV) nuclear antigen protein 1 (EBNA1, see AAY95856). EBNA1 is  
 CC utilized in a novel method for obtaining a eukaryotic cell that is  
 CC stably transfected with at least one episome. This method involves  
 CC transfecting a eukaryotic cell with: (1) a first episome comprising  
 CC an EBV origin of replication (oriP, see AAY50253), a gene encoding a  
 CC first protein whose expression results in cell death and a  
 CC selectable marker for eukaryotic cells; and (2) a second episome  
 CC comprising an EBV oriP and a gene encoding a second protein, where  
 CC expression of the second protein prohibits the occurrence of cell  
 CC death resulting from expression of the first protein to produce  
 CC doubly transfected cells which also express an antigen that  
 CC promotes retention of the episomes by the cells. The doubly  
 CC transfected cells are maintained under conditions in which the  
 CC first and second proteins and the selectable marker are expressed,  
 CC and the selective pressure specified by the marker is maintained.  
 CC Under these conditions, only cells containing both episomes live.  
 CC Preferably, EBNA1 is expressed from 1 of the episomes, and the  
 CC protein of interest from the other episome. Either or both  
 CC epitopes may further comprise a nucleic acid sequence encoding a  
 CC protein desired to be expressed in the cell (e.g. a therapeutic  
 CC protein); a nucleic acid encoding an RNA that is not intended to  
 CC be translated (e.g. a therapeutic RNA), or a DNA sequence used as  
 CC a tag for the cells. The method is applicable to cell culture or  
 CC intact organisms, for gene therapy. It allows the rapid  
 CC establishment of eukaryotic cells that stably and reliably express  
 CC a gene of interest, using a novel method of selection, and  
 CC maintenance of that selection without the need for exogenous  
 CC selection factors, such as antibiotics.

SO Sequence 1926 BP; 487 A; 352 C; 872 G; 215 T; 0 other;

Query Match 6.3%; Score 68.4; DB 21; Length 1926;  
 Best Local Similarity 44.0%; Pred. No. 7.3e-05;  
 Matches 288; Conservative 0; Mismatches 366; Indels 0; Gaps 0;

QY 406 GGGGGCGAAGAGAGTGGGACAGCAGACAGCTGGGAAAGAAAACCTAAGAGA 465  
 DB 305 GGGCAGAGAGAGAGAGGGGGCAGAGCAGAGAGGGGCGAGGGGCGAG 364  
 QY 466 CGCCCGTCGAAGAAGCGCGCTTGGAAACCGTAACTGAAGTGAAGAGAAG 525  
 DB 365 GAGGGCGAGAGAGAGAGGGGGCAGAGCAGAGAGAGGGGCGAGAGGGG 424  
 QY 526 AAAAGTTGACGAGAAAGAGAGCTTTCAGACTTCAAGATCCGAGCCGAGATGTTCCG 585

DB 425 CAGAGCAGAGAGAGGGGCGAGAGCAGAGAGGGGCGAGAGGCGAGAGGAG 484  
 QY 586 AAGGGCCAGCCCGTCGCGCCCTATAACACACGAGTTCTCATGATATACGACGAG 645  
 DB 485 GGGCAGAGAGGGGCGAGAGGGGCGAGAGCAGAGAGGGGCGAGAGGCGAGG 544  
 QY 646 GAGAGCCCGATCTCAAAACCGGCTGTATCTCAAGCGGGCCGCCAAATCCAGCAG 705  
 DB 545 GAGGGCGAGAGCAGAGAGAGGAGGAGGAGGGGCGAGAGGCGAGAGGAGG 604  
 QY 706 ACCAGCATGAGACTTCAAGAAAGAGGGGTTGAGAGGATGAGGGCGAGATGGGATG 765  
 DB 605 CAGAGCAGAGAGGGGCGAGAGGGGCGAGAGGCGAGAGAGGCGAGAGGAG 664  
 QY 766 GAGAGGAGCGGAGCGAGTTTCTGAGCGGGACTTCTCGAGACGTACGAGCGTACAC 825  
 DB 665 GAGGGCGAGAGGGGCGAGAGGGGCGAGAGCAGAGAGGGGCGAGAGGCGAGG 724  
 QY 826 ACGAGAGGCTTGCAGAACTGAGCAAGCAGAGGCTTATCAGAGAGTACTGAACTGAG 885  
 DB 725 CAGAGAGGGCGAGAGCAGAGAGGGGCGAGAGGCGAGAGCAGAGGCGAGAGG 784  
 QY 886 AAGTGCCTCTCGCATGAGAGCAGAAACACCGGCTGCGGCTGAGAGCAAGCGGCTG 945  
 DB 785 GAGCAGAGAGGGGCGAGAGGGGCGAGAGGAGGAGGCGAGAGGCGAGAGGAG 844  
 QY 946 GGTGGCAGCAGCGCGGTGTCGAGCTGAGAGCTGAGAGCTGCGCGCGCGAG 1005  
 DB 845 GAGGGCGAGAGGGGCGAGAGCAGAGAGGGGCGAGAGGGGCGAGAGGCGAGAG 904  
 QY 1006 AACCTTCAGCTCTGACCGAGAACTGACCGCGCAGCAGAGCGAGCGCGCG 1059  
 DB 905 GGGCAGAGCAGAGAGGGGCGAGAGGGGCGAGAGGCGAGAGAGGCGAGAGGAG 958

# RESULT 15

AAFB2902  
 ID AAF82902 standard; DNA; 1926 BP.

XX AAF82902;

AC 29-JUN-2001 (first entry)

DT EBV tethering protein EBNA1 encoding DNA.

XX Histone H1; tethering protein; LANA; gene therapy; multiple sclerosis;

KM Parkinson's disease; Huntington disease; diabetes; human herpesvirus 8;

KW EBV; latency-associated nuclear antigen; LANA; EBNA1; ds.

XX Epstein-barr virus.

OS Key Location/Qualifiers

FT CDS 1..1926

PN MO200125484-A2.

PD 12-APR-2001.

PF 29-SEP-2000; 2000MO-US26908.

PR 01-OCT-1999; 99US-0410399.

PA (UNM) UNIV MICHIGAN.

PI Robertson ES, Cotter MA;

DR WPI; 2001-281736/29.

PT P-PSDB; AAB62332.

A composition for use in gene therapy comprises an expression vector that includes a nucleic acid sequence encoding a nucleic acid binding

PT protein  
XX  
PS Disclosure; Fig 9C; 60bp; English.  
XX

CC The invention provides a composition comprising nucleic acid, histone H1  
CC protein and expression vector operably encoding a protein suitable  
CC for tethering the nucleic acid to the histone H1 protein, where the  
CC tethering protein is IANA. The composition is useful in aiding the  
CC retention of the viral DNA in the host cell. The viral vector encodes a  
CC protein suitable for tethering DNA to Histone H1. Methods for screening  
CC for compounds which are agonistic or antagonistic for the tethering of  
CC viral proteins to histone H1 and DNA binding sites are useful for  
CC developing the method of viral transfer. The composition has applications  
CC to gene therapy, including the treatment of multiple sclerosis,  
CC Parkinson's disease, Huntington disease and diabetes. The present  
CC sequence represents the nucleotide sequence of the Epstein-barr virus  
CC (EBV) tethering protein ENNA1.

XX  
SQ Sequence 1926 BP; 487 A; 352 C; 872 G; 215 T; 0 other;

Query Match 6.3%; Score 68.4; DB 22; Length 1926;  
Best Local Similarity 44.0%; Pred. No. 7.3e-05;  
Matches 288; Conservative 0; Mismatches 366; Indels 0; Gaps 0;

QY 406 GGGGGCGAAGAGAGTGGGACACAGCAGACAGCTGGGGAAGAAAACTAAGAGA 465  
Db 305 GGGCAGAGCAGAGAGAGGGGCGAGAGCAGAGAGGGGCGAGAGGGGCGAG 364  
QY 466 CGCCCTCCAAAGAACCGGCAATTGGAAACCTACTCAAGCTGAATGGGAAGAA 525  
Db 365 GAGGGCAGAGCAGAGAGAGGGGCGAGAGCAGAGAGGGGCGAGAGGGGCGAG 424  
QY 526 AAAAAGTTGACGAGAAACAGAGCTTCCAGCTTCAAGATCCGAGCCGAGATGTTCC 585  
Db 425 CAGAGCAGAGAGAGAGGGGCGAGAGCAGAGAGAGGGGCGAGAGCGAGAGAG 484  
QY 586 AAGGGCCAGCGGCTCGCCCTTAACACACAGCAGTTCCTCATGATGATCAACAG 645  
Db 485 GGGCAGAGAGGGGCGAGAGGGGCGAGAGCAGAGAGGGGCGAGAGCGAGAGGG 544  
QY 646 GAGGAGCCGATCTCAAAACCGGCTGTACTCAAGCGGCGCCCGCAATCCGAC 705  
Db 545 GAGGGCAGAGAGAGAGGGGCGAGAGGGGCGAGAGGGGCGAGAGCGAGAGGG 604  
QY 706 ACAGAGGATGACGACTTCATGAAAGAGGGGCTGAGAGGATGGGGCAGCGATG 765  
Db 605 CAGAGCAGAGAGGGGCGAGAGGGGCGAGAGGGGCGAGAGGGGCGAGAGAG 664  
QY 766 GAGAGGACGCGACGAGATTCTGACCGGACTTCTCGAGACGTAACGCGGTACAC 825  
Db 665 GAGGGGCGAGAGGGGCGAGAGGGGCGAGAGCGAGAGGGGCGAGAGGGGCGAG 724  
QY 826 ACCGAGAGCTTGCAAAATGAGCAAGCAGAGCTCATCAAGAGTACTTGAATGGAG 885  
Db 725 CAGAGAGGGCAGAGCAGAGGGGCGAGAGGGGCGAGAGCGAGAGGGGCGAG 784  
QY 886 AAGTGCTCTCGGCGCATGAGAGAGAGAACACCGCTGCGCTGAGAGCAAGCGCTG 945  
Db 785 GAGCAGAGAGGGGCGAGAGGGGCGAGAGCGAGAGAGGGGCGAGAGCGAGAG 844  
QY 946 GGTGCGACGACCGCGTGTGCGGAGCTGAGCTGAGCTGAGCGGCTGCGCGCAG 1005  
Db 845 GAGGGGCGAGAGGGGCGAGAGCGAGAGGGGCGAGAGGGGCGAGAGGGGCGAG 904  
QY 1006 AACTCTCAGCTGCTGACCGAAGAACTGCAACCGCAGCAGAGCGAGCGCTG 1059  
Db 905 GGGCAGAGCAGAGAGGGGCGAGAGGGGCGAGAGCGAGAGGGGCGAGAGCGAG 958

Search completed: February 5, 2004, 10:08:37  
Job time : 408 secs

THIS PAGE BLANK (USF10)



GenCore version 5.1.6  
Copyright (c) 1993 - 2004 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: February 5, 2004, 09:48:27 ; Search time 4185 Seconds

(without alignments)  
10557.323 Million cell updates/sec

Title: US-09-972-758A-1

Perfect score: 1080  
Sequence: 1 atggccgagccatcttctgc.....ttcccaagtttgagactag 1080

Scoring table: IDENTITY\_NUC  
Gapop 10.0, Gapext 1.0

Searched: 2888711 seqs, 2045481386 residues

Total number of hits satisfying chosen parameters: 5777422

Minimum DB seq length: 0  
Maximum DB seq length: 2000000000

Post-Processing: Minimum Match 0%  
Maximum Match 100%

Listing first 45 summaries

Database:

1: gb\_ba:\*  
2: gb\_hcg:\*  
3: gb\_in:\*  
4: gb\_om:\*  
5: gb\_ov:\*  
6: gb\_pac:\*  
7: gb\_ph:\*  
8: gb\_pl:\*  
9: gb\_pr:\*  
10: gb\_ro:\*  
11: gb\_sbs:\*  
12: gb\_sy:\*  
13: gb\_un:\*  
14: gb\_vl:\*  
15: em\_ba:\*  
16: em\_fun:\*  
17: em\_hum:\*  
18: em\_in:\*  
19: em\_mu:\*  
20: em\_om:\*  
21: em\_or:\*  
22: em\_ov:\*  
23: em\_pac:\*  
24: em\_ph:\*  
25: em\_pl:\*  
26: em\_ro:\*  
27: em\_sbs:\*  
28: em\_un:\*  
29: em\_vl:\*  
30: em\_hcg\_hum:\*  
31: em\_hcg\_huv:\*  
32: em\_hcg\_ocher:\*  
33: em\_hcg\_mus:\*  
34: em\_hcg\_pin:\*  
35: em\_hcg\_rod:\*  
36: em\_hcg\_mam:\*  
37: em\_hcg\_vrt:\*  
38: em\_sy:\*  
39: em\_hugo\_hum:\*  
40: em\_hugo\_mus:\*  
41: em\_hugo\_other:\*

score greater than or equal to the score of the result being printed,  
and is derived by analysis of the total score distribution.

#### SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	1076.8	99.7	2199	6 BD106410	BD106410 Secreted
2	1076.8	99.7	3624	9 AB021179	AB021179 Homo sapi
3	1076.8	99.7	35050	9 AC142472	AC142472 Homo sapi
4	1076.8	99.7	38849	9 AC138150	AC138150 Homo sapi
5	1076.8	99.7	100836	2 AC087288	AC087288 Homo sapi
6	1075.2	99.6	2178	6 BD160090	BD160090 Primer to
7	1075.2	99.6	2178	6 AC023624	AC023624 Homo sapi
8	1073.6	99.4	2086	9 BC006460	BC006460 Homo sapi
9	1052.8	97.5	142326	2 AC024047	AC024047 Homo sapi
10	840.8	77.9	3402	10 AY090614	AY090614 Mus muscu
11	840.8	77.9	111002	10 AL731805	AL731805 Mouse DNA
12	833.8	77.2	1488	10 BC022111	BC022111 Mus muscu
13	818	75.7	213625	2 AC120950	AC120950 Rattus no
14	471.2	43.6	111803	2 AC136772	AC136772 Rattus no
15	395.8	36.6	523	6 BD056159	BD056159 Secreted
16	355.2	32.9	461	6 AX31371	AX31371 Sequence
17	176.8	16.4	239464	2 AC103079	AC103079 Rattus no
18	176.8	16.4	267540	2 AC107153	AC107153 Rattus no
19	174.8	16.2	1238	9 BC025970	BC025970 Homo sapi
20	174.8	16.2	1330	6 AX714539	AX714539 Sequence
21	174.8	16.2	1330	9 AK056946	AK056946 Homo sapi
22	174.8	16.2	96875	9 AC008105	AC008105 Homo sapi
23	174.2	16.1	111603	2 AC136172	AC136172 Rattus no
24	173.2	16.0	2048	9 AK096785	AK096785 Homo sapi
25	172	15.9	223442	10 AL662804	AL662804 Mouse DNA
26	138	12.8	342	6 BD059439	BD059439 Secreted
27	110.8	10.3	225116	2 AC112538	AC112538 Rattus no
28	104.6	9.7	124088	10 AL929026	AL929026 Mouse DNA
29	104.4	9.7	240931	2 AC107097	AC107097 Rattus no
30	99.4	9.2	82400	10 AC090495	AC090495 Genomic s
31	96.4	8.9	171570	10 AL807807	AL807807 Mouse DNA
32	95.6	8.9	70748	2 AC137575	AC137575 Homo sapi
33	93.2	8.6	160539	2 AC109193	AC109193 Mus muscu
34	93	8.6	94058	2 AC141024	AC141024 Rattus no
35	91.8	8.5	76941	10 BX000698	BX000698 Mouse DNA
36	91.2	8.4	152435	10 AC121566	AC121566 Mus muscu
37	90.2	8.4	178757	10 AC122509	AC122509 Mus muscu
38	89.6	8.3	165592	10 AC122020	AC122020 Mus muscu
39	88.6	8.2	177100	2 AC112388	AC112388 Rattus no
40	88.6	8.2	246774	2 AC095078	AC095078 Rattus no
41	87.2	8.1	190944	2 AC113108	AC113108 Mus muscu
42	87.2	8.1	224198	2 AC106674	AC106674 Rattus no
43	87	8.1	205054	10 AC098712	AC098712 Mus muscu
44	86.6	8.0	201074	2 AC102097	AC102097 Mus muscu
45	86.2	8.0	178620	10 AC121865	AC121865 Mus muscu

#### ALIGNMENTS

RESULT 1  
LOCUS BD106410 2199 bp DNA linear PAT 18-SHP-2002  
DEFINITION Secreted proteins and polynucleotides encoding them.  
ACCESSION BD106410  
VERSION BD106410.1 GI:23201228  
KEYWORDS UP 2002503955-A/1.  
SOURCE Chlamydia sp.  
ORGANISM Chlamydia sp.  
REFERENCE 1 (bases 1 to 2199)  
Bacteria: Chlamydiae; Chlamydiales; Chlamydiaceae; Chlamydia.  
AUTHORS Jacobs,K., McCoy,J.M., Lavallie,E.R., Racie,L.A., Merberg,D.,  
Treacy,M., Spaulding,V. and Agostino,M.U.  
TITLE Secreted proteins and polynucleotides encoding them  
JOURNAL Patent: JP 2002503955-A 1 05-FEB-2002;

Pred. No. is the number of results predicted by chance to have a

COMMENT

GENETICS INSTITUTE INC  
 PN JP 2002503955-A/1  
 PD 05-FEB-2002  
 PF 20-MAR-1998 JP 1998545874  
 PR 21-MAR-1997 US 08/822167,19-MAR-1998 US 09/044466 PI  
 KENNETH JACOBS, JOHN M MCCOY, EDWARD R LAVALLIE, LISA A RACIE, PI  
 DAVID WEBBERG,  
 PI MAURICE TREACY, VIKKI SPAULDING, MICHAEL J AGOSTINO PC  
 C12N15/12, C07K14/47, A61K38/17  
 CC Strandedness: Doublet;  
 CC Topology: Linear;  
 FH Key Location/Qualifiers.

FEATURES

Source Location/Qualifiers  
 1. .2199  
 /organism="Chlamydia sp."  
 /mol\_type="genomic DNA"  
 /db\_xref="taxon:35827"

BASE COUNT 552 a 511 c 674 g 462 t

ORIGIN

Query Match 99.7%; Score 1076.8; DB 6; Length 2199;  
 Best Local Similarity 99.8%; Pred. No. 4.6e-169;  
 Matches 1078; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 1 ATGGCCGAGCCATTCTTGTCAGAATATCAACACCGCTTCAAACTAGCACTGTACAGGT 60  
 DB 707 ATGGCCGAGCCATTCTTGTCAGAATATCAACACCGCTTCAAACTAGCACTGTACAGGT 766  
 QY 61 GCTGCTCTCTCCAGGAAGCTGGAACCTGAGGCGCCCGGCGGAGAGCGGGGTG 120  
 DB 767 GCTGCTCTCTCCAGGAAGCTGGAACCTGAGGCGCCCGGCGGAGAGCGGGGTG 826  
 QY 121 CCCGAGAGAGCACTAGTGGCAATCGAGACGTTCCCGCAGTTGGTGGCCGTCGCGG 180  
 DB 827 CCCGAGAGAGCACTAGTGGCAATCGAGACGTTCCCGCAGTTGGTGGCCGTCGCGG 886  
 QY 181 CCGGAGGGGGAAGGAGCCTTGAATCCCAACCACTCTCTTGCAGACCCAGGCTGTCCA 240  
 DB 887 CCGGAGGGGGAAGGAGCCTTGAATCCCAACCACTCTCTTGCAGACCCAGGCTGTCCA 946  
 QY 241 GAATCTAGTGCCTGAGAGAGGGGAGAGGGGAGAGGGGAGAGGGGAGAGGGGAGAG 300  
 DB 947 GAATCTAGTGCCTGAGAGAGGGGAGAGGGGAGAGGGGAGAGGGGAGAGGGGAGAG 1006  
 QY 301 GGGGACTTCCCGCGCCGCGGAGAGAGTGAACCGACCGCCGAGCGGAGCTGTGCGCCAG 360  
 DB 1007 GGGGACTTCCCGCGCCGCGGAGAGAGTGAACCGACCGCCGAGCGGAGCTGTGCGCCAG 1066  
 QY 361 CTTTGTATATCTCCGAGCCAGTAACTTGGGGCTCTCTCCCGAGGGGCGAAGAGAG 420  
 DB 1067 CTTTGTATATCTCCGAGCCAGTAACTTGGGGCTCTCTCCCGAGGGGCGAAGAGAG 1126  
 QY 421 TGGGAGCAGAGAGAGAGAGCTGGGGAAGAAAAAATAGAGACCGCCGTCCAGAAAG 480  
 DB 1127 TGGGAGCAGAGAGAGAGAGCTGGGGAAGAAAAAATAGAGAGCCCGTCCAGAAAG 1186  
 QY 481 AAGCGGATTGGAACCGTACTCAAGCTGAATCTGGGAGAGAGAAAAAGTTGACAGAG 540  
 DB 1187 AAGCGGATTGGAACCGTACTCAAGCTGAATCTGGGAGAGAGAAAAAGTTGACAGAG 1246  
 QY 541 AAACAGAGCCTTGAAGCTTCAAGATCCGAGCCGAGATGTTGCCAAGGGCGACCGGTC 600  
 DB 1247 AAACAGAGCCTTGAAGCTTCAAGATCCGAGCCGAGATGTTGCCAAGGGCGACCGGTC 1306  
 QY 601 GGGCCCTATTAACCAAGGAGTCTCTCATGAGATGATCAACCAAGAGAGAGCGGATCTC 660  
 DB 1307 GGGCCCTATTAACCAAGGAGTCTCTCATGAGATGATCAACCAAGAGAGAGCGGATCTC 1366  
 QY 661 AAAACGGGCTGTACTCAAGCGGGCGCGCCAAATCCAGAGACCAAGCATGACGAC 720  
 DB 1367 AAAACGGGCTGTACTCAAGCGGGCGCGCCAAATCCAGAGACCAAGCATGACGAC 1426  
 QY 721 TTCATGAAGAGAGGGGCTGAAGAGATGGGGGAGCGATGGAGTGGAGAGGGAGCGGACG 780

DB 1427 TTCATGAAGAGAGGGGCTGAAGAGATGGGGGAGCGATGGAGTGGAGAGGGAGCGGACG 1486  
 QY 781 GAGTTCTCAGCGGAGCTTCTCGAGAGCTACGAGCGGTACCAACAGAGAGAGCTGCGAG 840  
 DB 1487 GAGTTCTCAGCGGAGCTTCTCGAGAGCTACGAGCGGTACCAACAGAGAGAGCTGCGAG 1546  
 QY 841 AACATGAGAGAGAGAGCTCATCAAGAGTACCTGGAACCTGAGAGAGTCCCTCTCGCGC 900  
 DB 1547 AACATGAGAGAGAGAGCTCATCAAGAGTACCTGGAACCTGAGAGAGTCCCTCTCGCGC 1606  
 QY 901 ATGAG 960  
 DB 1607 ATGAG 1666  
 QY 961 CGTGTGCGGAGAGCTGAGAGCTGAGAGCTGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1020  
 DB 1667 CGTGTGCGGAGAGCTGAGAGCTGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1726  
 QY 1021 ACCGAG 1080  
 DB 1727 ACCGAG 1786

RESULT 2  
 LOCUS AB021179 3624 bp mRNA linear PRI 05-DEC-2000  
 DEFINITION Homo sapiens mRNA for HEXIM1 protein, complete cds.  
 ACCESSION AB021179  
 VERSION AB021179.1 GI:4062855  
 KEYWORDS HEXIM1; HEXIM1 protein.  
 SOURCE Homo sapiens (human)  
 ORGANISM Homo sapiens  
 Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;  
 Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.  
 REFERENCE 1 (sites)  
 AUTHORS Kusunhara, M., Nagasaki, K., Kimura, K., Maas, N., Manabe, T.,  
 Ishikawa, S., Aikawa, M., Miyazaki, K. and Yamaguchi, K.  
 TITLE Cloning of hexamethylene-bis-acetamide-inducible transcript,  
 HEXIM1, in human vascular smooth muscle cells  
 JOURNAL Biomed. Res. 20, 273-279 (1999)  
 REFERENCE 2 (bases 1 to 3624)  
 AUTHORS Kusunhara, M.  
 TITLE Direct Submision  
 JOURNAL Submitted (14-DEC-1998) Masatoshi Kusunhara, National Defense  
 Medical College, The First Department of Internal Medicine; 3-2  
 Namiki, Tokorozawa, Saitama 359-8513, Japan  
 (E-mail: mkusun@nmdc.ac.jp, Tel: +81-42-995-1597,  
 Fax: +81-42-996-5200)

FEATURES

source Location/Qualifiers  
 1. .3624  
 /organism="Homo sapiens"  
 /mol\_type="mRNA"  
 /db\_xref="taxon:9606"  
 /cell\_type="coronary artery smooth muscle cell"  
 1. .3624  
 /gene="HEXIM1"  
 /gene="HEXIM1"  
 /codon\_start=1  
 /product="HEXIM1 protein"  
 /protein\_id="BAA36166.1"  
 /db\_xref="GI:4062855"  
 /translation="MAAPPLSYOHQROTNSCTGAAYVOELNRPPEGAERVRPEED  
 SEMGRAPFOLGRRPGRGEGSLSEQPPPLQAPRESSCLAREGKGNDDSSAGCD  
 PPPAEVPTPEAEILLOPCHDESKASKIGAPAGGEEBQOQROLGKRRRPSKK  
 KHMWRPYKLTWEEKKFKDEKQSLASRIEAMFPAKQPVAVYNTLOFLMDHDDEP  
 DLKTLVYKRAAKSGDDSDDFMEGCEEDGSDGMGDSSEFLQDFSEYERYHT  
 EQLQWAOELIKEYLELEKLSRMEDENNRLRLSESKRLGGDDAVRLELELDRLRA  
 ENLQLTENELHROERAFLSKFGD"

BASE COUNT 978 a 775 c 964 g 907 t

ORIGIN

Query Match 99.7%; Score 1076.8; DB 9; Length 3624;  
 Best Local Similarity 99.8%; Pred. No. 4.3e-169;  
 Matches 1078; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

```

QY 1 ATGGCCGACCCATTCTTTGATCAATATCAACACACCTTCAAACTGACATCTGTACAGT 60
DB 690 ATGGCCGACCCATTCTTTGATCAATATCAACACACCTTCAAACTGACATCTGTACAGT 749
QY 61 GGTGCTGCTGTCAGAGAGAGCTGAACCTTGAGCGGCCCGCCAGCGCGAGAGAGCGGGT 120
DB 750 GGTGCTGCTGTCAGAGAGAGCTGAACCTTGAGCGGCCCGCCAGCGCGAGAGAGCGGGT 809
QY 121 CCGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 180
DB 810 CCGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 869
QY 181 CCGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 240
DB 870 CCGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 929
QY 241 GAATCTAGCTGCTGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 300
DB 930 GAATCTAGCTGCTGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 989
QY 301 GGGGACTTCCCGCCCGCGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 360
DB 990 GGGGACTTCCCGCCCGCGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1049
QY 361 CTTTGTCTGATCTCCGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 420
DB 1050 CTTTGTCTGATCTCCGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1109
QY 421 TGGGGGACAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 480
DB 1110 TGGGGGACAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1169
QY 481 AAGCGGCTTGGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 540
DB 1170 AAGCGGCTTGGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1229
QY 541 AAGCGGCTTGGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 600
DB 1230 AAGCGGCTTGGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1289
QY 601 GGGCCCTATTAACCAACGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 660
DB 1290 GGGCCCTATTAACCAACGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1349
QY 661 AAAAAAGGCGCTGTAATCAAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 720
DB 1350 AAAAAAGGCGCTGTAATCAAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1409
QY 721 TTCTATGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 780
DB 1410 TTCTATGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1469
QY 781 GAGTTTCTGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 840
DB 1470 GAGTTTCTGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1529
QY 841 AACATGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 900
DB 1530 AACATGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1589
QY 901 ATGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 960
DB 1590 ATGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1649
QY 961 CGGTGCGGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1020
DB 1650 CGGTGCGGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1709
QY 1021 ACCGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1080
  
```

Db 1710 ACCGAG 1769

## RESULT 3

AC142472 35050 bp DNA linear PRI 04-APR-2003

LOCUS Homo sapiens chromosome 17, clone XHfos-8119G3, complete sequence.

DEFINITION AC142472

AC142472.1 GI:29540490

KEYWORDS HTG.

SOURCE Homo sapiens (human)

ORGANISM Homo sapiens

Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;

Mammalia; Eutheria; Primates; Catarrhini; Hominoidea; Homo.

1 (bases 1 to 35050)

2 (bases 1 to 35050)

REFERENCE Birren, B., Nusbaum, C., Lander, E., Aboueleil, A., Allen, N.,

Anderson, S., Arachchi, H.M., Barna, N., Bastien, V., Bloom, T.,

Boguslavsky, L., Boukhgalter, B., Camarata, J., Chang, J., Choepel, Y.,

Collins, A., Cook, A., Cooke, P., Corum, B., Dearellano, K.,

Diaz, J.S., Dodge, S., Dooley, K., Dorris, L., Erickson, J., Faro, S.,

Ferreira, P., Fitzgerald, M., Gage, D., Galagan, J., Gardyna, S.,

Graham, L., Grand-Pierre, N., Hafez, N., Hagopian, D., Hagos, B.,

Hall, J., Horton, L., Hulme, W., Iliev, I., Johnson, R., Jones, C.,

Kamat, A., Karatas, A., Kelle, C., Landers, T., Levine, R.,

Linblad, T., Liu, G., Lui, A., Mabbitt, R., MacLean, C.,

Macdonald, P., Major, J., Manning, J., Matthews, C., McCarthy, M.,

Meldrum, J., Meneus, L., Mihova, T., Mieng, V., Murphy, T., Naylor, J.,

Nguyen, C., Nicol, R., Norbu, C., O'Connor, T., O'Donnell, P.,

O'Neil, D., Oliver, J., Peterson, K., Phunhang, P., Pierre, N.,

Rachupka, A., Ramsamy, U., Raymond, C., Retta, R., Rise, C., Rogov, P.,

Roman, J., Schauer, S., Schupbach, R., Seaman, S., Severy, P., Smith, C.,

Spencer, B., Strange-Thomann, N., Stojanovic, N., Stubbs, M.,

Talimas, J., Testaf, S., Theodore, J., Topham, K., Travers, B.,

Vasilev, H., Venkataraman, V.S., Viel, R., Vo, A., Wilson, B., Wu, X.,

Wyman, D., Young, G., Zainoun, J., Zembek, L., Zimmer, A. and Zody, M.

Direct Submission

Submitted (04-APR-2003) Whitehead Institute/MIT Center for Genome

Research, 320 Charles Street, Cambridge, MA 02141, USA

All repeats were identified using RepeatMasker:

Smit, A.F.A. & Green, P. (1996-1997)

http://ftp.genome.washington.edu/RM/RepeatMasker.html

----- Genome Center

Center: Whitehead Institute/ MIT Center for Genome Research

Web site: http://www-seq.wi.mit.edu

Contact: sequence\_submissions@genome.wi.mit.edu

----- Project Information

Center project name: L29477

Center clone name: 8119\_G\_3

-----

Only the last 35.1 kilobases of this clone are being submitted.

The remainder overlaps accession number AC008105 (WICR project

L890).

----- Location/Qualifiers

1..35050

/organism="Homo sapiens"

/mol\_type="genomic DNA"

/db\_xref="taxon:9606"

/chromosome="17"

/map="17"

/clone="XHfos-8119G3"

/clone\_lib="Human Female Fosmid Library"

14..320

/rpt\_family="AluSp"

1947..1983

/rpt\_family="AT-rich"

repeat\_region

repeat\_region

repeat\_region

```

repeat_region /rpt_family="(A)n"
3993. .4099 /rpt_family="C-rich"
repeat_region 5028. .5069 /rpt_family="G-rich"
complement(5605. .5818)
repeat_region /rpt_family="MIR"
complement(6064. .6196)
repeat_region /rpt_family="L2"
complement(6713. .6862)
repeat_region /rpt_family="MIR"
6863. .7166 /rpt_family="AluSc"
repeat_region complement(7167. .7225)
/rpt_family="MIR"
8048. .8104 /rpt_family="(CCCCG)n"
repeat_region 8113. .8149 /rpt_family="GC-rich"
/rpt_family="GC-rich"
9472. .9732 /rpt_family="LMC4"
9733. .10036 /rpt_family="AluSg1"
10037. .10091 /rpt_family="LMC4"
/rpt_family="LMC4"
complement(10169. .10477)
repeat_region /rpt_family="AluSx"
10480. .10739 /rpt_family="LMC4"
10750. .10906 /rpt_family="LMC5"
/rpt_family="LMC5"
complement(10907. .11182)
repeat_region /rpt_family="AluJo"
11183. .11552 /rpt_family="LMC5"
complement(11556. .11857)
repeat_region /rpt_family="AluSx"
complement(11913. .12044)
/rpt_family="FLAM_C"
12108. .12240 /rpt_family="FLAM_C"
12296. .12374 /rpt_family="FLAM_C"
repeat_region /rpt_family="AluS"
12353. .12373 /note="<30 qual SNGI region"
12374 /note="unresolved sequence region:
100 bp of sequence missing here"
unreue
repeat_region 12375. .12673 /rpt_family="AluSc"
13059. .13232 /rpt_family="MIR"
complement(14041. .14336)
repeat_region /rpt_family="AluSx"
complement(14518. .14598)
/rpt_family="MIR"
17580. .17884 /rpt_family="AluSx"
17893. .18028 /rpt_family="LMMA3"
/rpt_family="LMMA3"
complement(18032. .18350)
repeat_region /rpt_family="AluY"
18360. .18383 /rpt_family="AT-rich"
complement(18387. .18515)
/rpt_family="FLAM_C"
18533. .18707 /rpt_family="FLAM"
18708. .18877 /rpt_family="LMMA4"
19015. .19324 /rpt_family="AluYas"
repeat_region /rpt_family="AluYas"
20221. .20288

```

```

repeat_region /rpt_family="(CCCCG)n"
20403. .20429 /rpt_family="GC-rich"
repeat_region complement(21264. .21365)
/rpt_family="MIR"
complement(21662. .21734)
repeat_region /rpt_family="L2"
complement(22401. .22550)
/rpt_family="L2"
complement(23139. .23338)
/rpt_family="MIR"
complement(23346. .23673)
/rpt_family="AluJo"
complement(23704. .23903)
/rpt_family="MIR"
complement(23996. .24125)
/rpt_family="MIR53"
24057. .24138 /rpt_family="MIR53"
25014. .25095 /rpt_family="(CA)n"
25274. .25344 /rpt_family="MIR3"
complement(25687. .25742)
/rpt_family="L2"
complement(25847. .25917)
/rpt_family="LMB8"
25918. .26185 /rpt_family="AluJo"
complement(26186. .26209)
/rpt_family="LMB8"
26243. .26497 /rpt_family="AluYc"
26500. .26799 /rpt_family="AluY"
complement(26800. .27209)
/rpt_family="LMB4"
complement(27220. .27528)
/rpt_family="AluSx"
complement(27533. .27592)
/rpt_family="AluSg/x"
complement(27597. .27800)
/rpt_family="AluSx"
complement(27805. .27916)
/rpt_family="AluSg/x"
complement(27917. .27963)
/rpt_family="LMB4"
28143. .29261 /rpt_family="MIR5B"
complement(29547. .29628)
/rpt_family="L2"
complement(30172. .30304)
/rpt_family="MIR"
30327. .30450 /rpt_family="MIR"
complement(30485. .30611)
/rpt_family="AluSg/x"
30675. .30825 /rpt_family="L13b"
complement(30871. .31136)
repeat_region

```

Query Match 99.7%; Score 1076.8; DB 9; Length 35050;  
 Best Local Similarity 99.8%; Pred. No. 3.1e-169;  
 Matches 1078; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

```

QY 1 ATGGCCGAGCCATTCTTTCAGAAATTCACACCAAGCTCAAACTAGCAACTGTACAGGT 60
DB 3397 ATGGCCGAGCCATTCTTTCAGAAATTCACACCAAGCTCAAACTAGCAACTGTACAGGT 3338
QY 61 GCTGCTGCTGTCAGGAAGAGCTGAACCTGAGCGCCGCCAGGCGGAGGAGCGGAGT 120
DB 3337 GCTGCTGCTGTCAGGAAGAGCTGAACCTGAGCGCCGCCAGGCGGAGGAGCGGAGT 3278

```

QY	121	CCCCAGGAGGACAGTAGGTGGCAATCCAGAGACGTTCCCCCAGTTGAGTGAGCGTC	CGGG	180
Db	3277	CCCCAGGAGGACAGTAGGTGGCAATCCAGAGACGTTCCCCCAGTTGAGTGAGCGTC	CGGG	3218
QY	181	CCGAGGGGGAGAGGAGGAGCTTGGAAATCCCAACCACTCCCTTGACAGCCCAAGCCCTGTCCA	240	
Db	3217	CCGAGGGGGAGAGGAGGAGCTTGGAAATCCCAACCACTCCCTTGACAGCCCAAGCCCTGTCCA	3158	
QY	241	GAATCTAGCTGCTTGAGAGAGGGCGAGAGAGGGCCAGAAATGGGGACAGACTCGTCCGCTGGC	300	
Db	3157	GAATCTAGCTGCTTGAGAGAGGGCGAGAGAGGGCCAGAAATGGGGACAGACTCGTCCGCTGGC	3098	
QY	301	GGGCACTTCCGCGCGCGCGAGAAAGTGGAAACCGACGCCCGAGGGCGAGCTCTGCCAC	360	
Db	3097	GGGCACTTCCGCGCGCGCGAGAAAGTGGAAACCGACGCCCGAGGGCGAGCTCTGCCAC	3038	
QY	361	CCTTGATGATGACTCCGAGGCCAGTAAATGTTGGGGCTCCTGCGCAGAGGGGGAGAGAGAG	420	
Db	3037	CCTTGATGATGACTCCGAGGCCAGTAAATGTTGGGGCTCCTGCGCAGAGGGGGAGAGAGAG	2978	
QY	421	TGGGGACAGCAGCAGAGACAGCTGCGGGAGAAAAAATTAAGAGACGCCGCTCAAGAG	480	
Db	2977	TGGGGACAGCAGCAGAGACAGCTGCGGGAGAAAAAATTAAGAGACGCCGCTCAAGAG	2918	
QY	481	AAGGCGCATTTGGAAACCGTACTTAACAGCTGAACTTGGGAAGAGAAAGAAATTTGACAG	540	
Db	2917	AAGGCGCATTTGGAAACCGTACTTAACAGCTGAACTTGGGAAGAGAAAGAAATTTGACAG	2858	
QY	541	AAACAGAGCCTTCGAGCTTCAAGAGTCCGAGCCGAGATGTTTCGCAAGAGGCGCAGCCGGT	600	
Db	2857	AAACAGAGCCTTCGAGCTTCAAGAGTCCGAGCCGAGATGTTTCGCAAGAGGCGCAGCCGGT	2798	
QY	601	GCGCCCTATAACACACACGACAGTTCTCTCATGAGTATCAACAGACAGAGAGCCGAGATTC	660	
Db	2797	GCGCCCTATAACACACACGACAGTTCTCTCATGAGTATCAACAGACAGAGAGCCGAGATTC	2738	
QY	661	AAAAACGGGCTGTACTCCAACGGGGCGCGCCCAATTCGACAGCACTCAGAGATGACGAC	720	
Db	2737	AAAAACGGGCTGTACTCCAACGGGGCGCGCCCAATTCGACAGCACTCAGAGATGACGAC	2678	
QY	721	TTTCATGGAAGAAAGGGGGGTGAGAGGATGGGGGCAAGCGATGGGATGGGAGGGGACGGCAGC	780	
Db	2677	TTTCATGGAAGAAAGGGGGGTGAGAGGATGGGGGCAAGCGATGGGATGGGAGGGGACGGCAGC	2618	
QY	781	GAGTTTCTGCAGCGGGACCTTCTCGGAGACGTATCAGACGGTATCCACAACGAGAGCCTGCGAG	840	
Db	2617	GAGTTTCTGCAGCGGGACCTTCTCGGAGACGTATCAGAGCGGTATCCACAACGAGAGCCTGCGAG	2558	
QY	841	AACATGAGCAACAGCAGAGACTCATCAGAGATTAAGTGAATCTTGAGAGAAATGCTCTCGCGC	900	
Db	2557	AACATGAGCAACAGCAGAGACTCATCAGAGATTAAGTGAATCTTGAGAGAAATGCTCTCGCGC	2498	
QY	901	ATGAGAGACGAGAACACCGGCTCGGTGAGAGCAACGGGCTGGGTGGGAGCGA	960	
Db	2497	ATGAGAGACGAGAACACCGGCTCGGTGAGAGCAACGGGCTGGGTGGGAGCGA	2438	
QY	961	CGTGTGCGGAGCTGAGAGCTGAGAGCTGGACCGGCTGCGCGCGAGAACCTTCAGCTGCTG	1020	
Db	2437	CGTGTGCGGAGCTGAGAGCTGAGAGCTGGACCGGCTGCGCGCGAGAACCTTCAGCTGCTG	2378	
QY	1021	ACCGAGAAACGAATGACACCGGACAGAGAGCGAGCGCGCTTTCCAAATTTGAGACTAG	1080	
Db	2377	ACCGAGAAACGAATGACACCGGACAGAGAGCGAGCGCGCTTTCCAAATTTGAGACTAG	2318	

ORGANISM  
Homo sapiens  
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.  
REFERENCE  
1 (bases 1 to 38849)  
AUTHORS  
Birtten,B., Nusbaum,C. and Lander,E.  
TITLE  
Homo sapiens chromosome 17, clone RP13-890H12  
JOURNAL  
Unpublished  
REFERENCE  
2 (bases 1 to 38849)  
AUTHORS  
Birtten,B., Nusbaum,C., Lander,E., Ali,A., Allen,N., Anderson,S.,  
Barna,N., Bastien,V., Bloom,T., Boguslavsky,L., Boukhalter,B.,  
Camarata,J., Chang,J., Chazaro,B., Choepel,Y., Collymore,A.,  
Cook,A., Cooke,P., Dearellano,K., Dewar,K., Diaz,J.S., Dodge,S.,  
Faro,S., Ferreira,P., Fitzgerald,M., Gage,D., Galagan,J.,  
Gardyna,S., Gordon,L., Graham,L., Grand-Pierre,N., Hafez,N.,  
Hagos,B., Horton,L., Hulme,W., Iliev,I., Johnson,R., Jones,C.,  
Kamat,A., Karatas,A., Kells,C., Landers,T., Levine,R.,  
Lindblad-Toh,K., Liu,G., Maclean,C., Macdonald,P., Major,J.,  
Matthews,C., McCarthy,M., Meldrum,J., Meneus,L., Mihova,T.,  
Mienga,V., Murphy,T., Naylor,J., Nguyen,C., Nicol,R., Norbu,C.,  
Norman,C.H., O'Connor,T., O'Donnell,P., O'Neil,D., Oliver,J.,  
Peterson,K., Phunhngang,P., Pierre,N., Raymond,C., Retta,R.,  
Rise,C., Rogov,P., Roman,J., Roy,A., Schauer,S., Schupback,R.,  
Seaman,S., Severy,P., Smith,C., Spencer,B., Stange-Thomann,N.,  
Stojanovic,N., Talamas,J., Teafaye,S., Theodore,J., Topham,K.,  
Travers,M., Vassiliev,H., Viel,R., Vo,A., Wilson,B., Wu,X.,  
Wyman,D., Young,G., Zainoun,J., Zempek,L., Zimmer,A. and Zody,M.  
JOURNAL  
Direct Submission  
TITLE  
Submitted (18-DEC-2002) Whitehead Institute/MIT Center for Genome  
REFERENCE  
Research, 320 Charles Street, Cambridge, MA 02141, USA  
AUTHORS  
3 (bases 1 to 38849)  
Birtten,B., Nusbaum,C., Lander,E., Abouelleil,A., Allen,N.,  
Anderson,S., Arachchi,H.M., Barna,N., Bastien,V., Bloom,T.,  
Boguslavsky,L., Boukhalter,B., Camarata,R., Chang,J., Choepel,Y.,  
Collymore,A., Cook,A., Cooke,P., Corum,B., Dearellano,K.,  
Diaz,J.S., Dodge,S., Dooley,K., Dorris,L., Erickson,J., Faro,S.,  
Ferreira,P., Fitzgerald,M., Gage,D., Galagan,J., Gardyna,S.,  
Graham,L., Grand-Pierre,N., Hafez,N., Hagopian,D., Hagos,B.,  
Hall,J., Horton,L., Hulme,W., Iliev,I., Johnson,R., Jones,C.,  
Kamat,A., Karatas,A., Kells,C., Landers,T., Levine,R.,  
Lindblad-Toh,K., Liu,G., Lui,A., Mabbitt,R., Maclean,C.,  
Macdonald,P., Major,J., Manning,J., Matthews,C., McCarthy,M.,  
Meldrum,J., Meneus,L., Mihova,T., Mienga,V., Murphy,T., Naylor,J.,  
Nguyen,C., Nicol,R., Norbu,C., O'Connor,T., O'Donnell,P.,  
O'Neil,D., Oliver,J., Peterson,K., Phunhngang,P., Pierre,N.,  
Rachpaka,A., Ramasamy,U., Raymond,C., Retta,R., Rise,C., Rogov,P.,  
Roman,J., Schauer,S., Schupback,R., Seaman,S., Severy,P., Smith,C.,  
Spencer,B., Stange-Thomann,N., Stojanovic,N., Stubbs,M.,  
Talamas,J., Teafaye,S., Theodore,J., Topham,K., Travers,M.,  
Vassiliev,H., Venkataraman,V.S., Viel,R., Vo,A., Wilson,B., Wu,X.,  
Wyman,D., Young,G., Zainoun,J., Zempek,L., Zimmer,A. and Zody,M.  
JOURNAL  
Direct Submission  
TITLE  
Submitted (31-JAN-2003) Whitehead Institute/MIT Center for Genome  
REFERENCE  
Research, 320 Charles Street, Cambridge, MA 02141, USA  
AUTHORS  
4 (bases 1 to 38849)  
Birtten,B., Nusbaum,C., Lander,E., Abouelleil,A., Allen,N.,  
Anderson,S., Arachchi,H.M., Barna,N., Bastien,V., Bloom,T.,  
Boguslavsky,L., Boukhalter,B., Camarata,R., Chang,J., Choepel,Y.,  
Collymore,A., Cook,A., Cooke,P., Corum,B., Dearellano,K.,  
Diaz,J.S., Dodge,S., Dooley,K., Dorris,L., Erickson,J., Faro,S.,  
Ferreira,P., Fitzgerald,M., Gage,D., Galagan,J., Gardyna,S.,  
Graham,L., Grand-Pierre,N., Hafez,N., Hagopian,D., Hagos,B.,  
Hall,J., Horton,L., Hulme,W., Iliev,I., Johnson,R., Jones,C.,  
Kamat,A., Karatas,A., Kells,C., Landers,T., Levine,R.,  
Lindblad-Toh,K., Liu,G., Lui,A., Mabbitt,R., Maclean,C.,  
Macdonald,P., Major,J., Manning,J., Matthews,C., McCarthy,M.,  
Meldrum,J., Meneus,L., Mihova,T., Mienga,V., Murphy,T., Naylor,J.,  
Nguyen,C., Nicol,R., Norbu,C., O'Connor,T., O'Donnell,P.,  
O'Neil,D., Oliver,J., Peterson,K., Phunhngang,P., Pierre,N.,  
Rachpaka,A., Ramasamy,U., Raymond,C., Retta,R., Rise,C., Rogov,P.,  
Roman,J., Schauer,S., Schupback,R., Seaman,S., Severy,P., Smith,C.,  
Spencer,B., Stange-Thomann,N., Stojanovic,N., Stubbs,M.,  
Talamas,J., Teafaye,S., Theodore,J., Topham,K., Travers,M.,  
Vassiliev,H., Venkataraman,V.S., Viel,R., Vo,A., Wilson,B., Wu,X.,  
Wyman,D., Young,G., Zainoun,J., Zempek,L., Zimmer,A. and Zody,M.

TITLE Wyman,D., Young,G., Zainoun,J., Zembek,L., Zimmer,A. and Zody,M.  
Journal Direct Submission  
Submitted (14-FEB-2003) Whitehead Institute/MIT Center for Genome  
Research, 320 Charles Street, Cambridge, MA 02141, USA

## COMMENT

On Feb 14, 2003 this sequence version replaced gi:28172166.  
All repeats were identified using RepeatMasker:  
Smit, A.P.A. & Green, P. (1996-1997)  
http://ftp.genome.washington.edu/RM/RepeatMasker.html

Center: Whitehead Institute/ MIT Center for Genome Research  
Genome Center

Center code: WIBR  
Web site: http://www-seq.wi.mit.edu

Contact: sequence\_submissions@genome.wi.mit.edu  
Project Information

Center project name: L28731  
Center clone name: 890\_H\_12

## FEATURES

## source

1..38849 Location/Qualifiers  
/organism="Homo sapiens"  
/mol\_type="Genomic DNA"  
/db\_xref="taxon:9606"  
/chromosome="17"  
/map="17"  
/clone="RP13-890H12"  
/clone\_1b="RP1-13 Human Female BAC"  
84..385  
/rpt\_family="AluSx"  
complement(389..758)  
/rpt\_family="L1MC5"  
759..1034  
/rpt\_family="AluJo"  
complement(1035..1191)  
/rpt\_family="L1MC5"  
complement(1202..1461)  
/rpt\_family="L1MC4"  
1464..1772  
/rpt\_family="AluSx"  
complement(1850..1904)  
/rpt\_family="L1MC4"  
complement(1905..2208)  
/rpt\_family="AluSg1"  
complement(2209..2469)  
/rpt\_family="L1MC4"  
3792..3828  
/rpt\_family="GC\_rich"  
3837..3893  
/rpt\_family="CGGGG)n"  
4716..4774  
/rpt\_family="MIR"  
complement(4775..5078)  
/rpt\_family="AluSc"  
5079..5204  
/rpt\_family="MIR"  
5745..5877  
/rpt\_family="L2"  
6123..6336  
/rpt\_family="MIR3"  
6872..6913  
/rpt\_family="C-rich"  
7842..7948  
/rpt\_family="G-rich"  
8448..8472  
/rpt\_family="(T)n"  
9962..9998  
/rpt\_family="AT\_rich"  
complement(11625..11932)  
/rpt\_family="AluSp"  
12103..12389  
/rpt\_family="AluSp"  
complement(12567..12862)  
/rpt\_family="AluSx"  
complement(12863..13169)

repeat\_region /rpt\_family="AluJb"  
complement(13209..13491)  
repeat\_region /rpt\_family="AluY"  
13566..13589  
repeat\_region /rpt\_family="AT\_rich"  
13595..13903  
repeat\_region /rpt\_family="AluSp"  
13904..14058  
repeat\_region /rpt\_family="AluSx"  
14039..14352  
repeat\_region /rpt\_family="AluY"  
14353..14522  
repeat\_region /rpt\_family="AluSx"  
14558..14685  
repeat\_region /rpt\_family="FLAM\_C"  
complement(14704..14854)  
/rpt\_family="FRAM"  
repeat\_region 15205..15307  
/rpt\_family="AluJb"  
15359..15633  
/rpt\_family="AluSx"  
16074..16377  
/rpt\_family="AluSc"  
16380..16558  
/rpt\_family="AluSg/x"  
16743..16877  
/rpt\_family="AluSx"  
16878..17181  
/rpt\_family="AluSp"  
17182..17348  
/rpt\_family="AluSx"  
complement(18100..18399)  
/rpt\_family="AluSc"  
complement(18948..19238)  
/rpt\_family="AluSx"  
22406..22707  
/rpt\_family="AluSx"  
complement(22708..22845)  
/rpt\_family="MIR"  
22960..23228  
/rpt\_family="AluSx"  
23231..23524  
/rpt\_family="AluSg"  
23532..23838  
/rpt\_family="AluSx"  
23843..24006  
/rpt\_family="AluSg/x"  
25108..25151  
repeat\_region

Query Match 99.7%; Score 1076.8; DB 9; Length 38849;

Best Local Similarity 99.8%; Pred. No. 3.1e-169;

Matches 1078; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 1 ATGGCCGAGCATTCTTTCAGAAATTCACACACCAAGCTCAAACTGTAAGT 60  
DB 8544 ATGGCCGAGCATTCTTTCAGAAATTCACACACCAAGCTCAAACTGTAAGT 603  
QY 61 GCTGCTGCTGTCCAGAAAGCTGAACTCTGAGCGCCCGGAGGCGGAGT 120  
DB 8604 GCTGCTGCTGTCCAGAAAGCTGAACTCTGAGCGCCCGGAGGCGGAGT 8663  
QY 121 CCGAGAGAGACAGTAGTGAGCAATGAGAGCGTTCCCACTTGAGTGGCCGCGG 180  
DB 8664 CCGAGAGAGACAGTAGTGAGCAATGAGAGCGTTCCCACTTGAGTGGCCGCGG 8723  
QY 181 CCGAGAGAGAGAGAGAGCTGGAATCCCAACCACTCTTTCAGACCCAGGCTTCCA 240  
DB 8724 CCGAGAGAGAGAGAGAGCTGGAATCCCAACCACTCTTTCAGACCCAGGCTTCCA 8783  
QY 241 GAATCTAGCTGCTGAGAGAGGCGGAGAGGCGGAGAGAGAGAGAGAGAGAGAG 300  
DB 8784 GAATCTAGCTGCTGAGAGAGGCGGAGAGGCGGAGAGAGAGAGAGAGAGAGAG 8843

```

QY 301 GGGGACTTCCCGCCGCGCAGAGTGGAAACCAAGCCGCGGAGGCGGAGCTGCTGCCAG 360
DB 8844 GGGGACTTCCCGCCGCGCAGAGTGGAAACCAAGCCGCGGAGGCGGAGCTGCTGCCAG 8903
QY 361 CCTGTGATGACCTCCGAGGCGCACTAAGTGGGGGCTCTCGCGCAGGCGGCGCAAGAGAG 420
DB 8904 CCTGTGATGACCTCCGAGGCGCACTAAGTGGGGGCTCTCGCGCAGGCGGCGCAAGAGAG 8963
QY 421 TGGGAGCAGAGCAGAGCAGCTGGGGAGAAAAAATATAGAGAGCGCCCTCCAGAAAG 480
DB 8964 TGGGAGCAGAGCAGAGCAGCTGGGGAGAAAAAATATAGAGAGCGCCCTCCAGAAAG 9023
QY 481 AAGCGGCAATGGAAACCGTACTCAAGCTGAGTGGAGAGAGAAAAAGTTTCAAGAG 540
DB 9024 AAGCGGCAATGGAAACCGTACTCAAGCTGAGTGGAGAGAGAAAAAGTTTCAAGAG 9083
QY 541 AAGCAGAGCTTGGAGCTTCAAGGATCCGAGCCGAGATGTTGCCCAAGGGCCAGCCGCTC 600
DB 9084 AAGCAGAGCTTGGAGCTTCAAGGATCCGAGCCGAGATGTTGCCCAAGGGCCAGCCGCTC 9143
QY 601 GCGCCCTATTAACACACAGCAGTTCCTCATGATGATCAACAGCAGAGAGCCGATCTC 660
DB 9144 GCGCCCTATTAACACACAGCAGTTCCTCATGATGATCAACAGCAGAGAGCCGATCTC 9203
QY 661 AAAACCGGCTGTACTTCAAGCGGCGCGCCGCAAAATCCGACGACACCAAGCATGACAG 720
DB 9204 AAAACCGGCTGTACTTCAAGCGGCGCGCCGCAAAATCCGACGACACCAAGCATGACAG 9263
QY 721 TTATGAGAGAGAGGGGCTGAGAGAGATGGGGGACGATGGGATGGGAGGGGACGGCAGC 780
DB 9264 TTATGAGAGAGAGGGGCTGAGAGAGATGGGGGACGATGGGATGGGAGGGGACGGCAGC 9323
QY 781 GAGTTTCTGAGGAGGGGACTTCTCGAGAGCGTACGAGCGGATCAACACGAGAGCCTGACAG 840
DB 9324 GAGTTTCTGAGGAGGGGACTTCTCGAGAGCGTACGAGCGGATCAACACGAGAGCCTGACAG 9383
QY 841 AACATGAGCAGAGAGAGAGCTCATCAAGAGTACTTGGAAAGTGGAGAGTCTCTCGCGC 900
DB 9384 AACATGAGCAGAGAGAGAGCTCATCAAGAGTACTTGGAAAGTGGAGAGTCTCTCGCGC 9443
QY 901 ATGAGAGAGCAGAGAAACACCGGCTGCGGCTGAGAGACAAAGCGCTGGGTGCGACGACGG 960
DB 9444 ATGAGAGAGCAGAGAAACACCGGCTGCGGCTGAGAGACAAAGCGCTGGGTGCGACGACGG 9503
QY 961 CGGTGCGGAGAGCTGAGAGCTGAGAGCTGCGGCTGCGGCGGAGAACTCCAGCTGCTG 1020
DB 9504 CGGTGCGGAGAGCTGAGAGCTGAGAGCTGCGGCTGCGGCGGAGAACTCCAGCTGCTG 9563
QY 1021 ACCGAGAACGAACTGACCGGCGAGCAGAGCGCGCTTCCAAAGTTTGGAGACTAG 1080
DB 9564 ACCGAGAACGAACTGACCGGCGAGCAGAGCGCGCTTCCAAAGTTTGGAGACTAG 9623

```

```

RESULT 5
AC087298/c 100836 bp DNA linear HTG 20-APR-2002
LOCUS Homo sapiens chromosome 17 clone RP11-40363 map 17, WORKING DRAFT
DEFINITION
SEQUENCE 15 ordered pieces.
AC087298 AC087298 GI:20219131
AC087298.8 HTGS PHASE2; HTGS_DRAFT; HTGS_FULLTOP.
VERSION HTGS PHASE2; HTGS_DRAFT; HTGS_FULLTOP.
KEYWORDS Homo sapiens (human)
SOURCE Homo sapiens
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
REFERENCE 1 (bases 1 to 100836)
AUTHORS Birren,B., Linton,L., Nusbaum,C., Lander,E., Allen,N., Anderson,S.,
Barna,N., Baskien,V., Boguslavsky,L., Boukhgalter,B., Brown,A.,
Camata,J., Campolano,A., Choepel,Y., Colangelo,M., Collins,S.,

```

# TITLE JOURNAL REFERENCE AUTHORS

## TITLE JOURNAL COMMENT

Collimore, A., Cooke, P., Dearellano, K., Dewar, K., Diaz, J.S.,  
Dodge, S., Faro, S., Ferreira, P., Fitzhugh, W., Gage, D., Galagan, J.,  
Gardyna, S., Ginde, S., Goyette, M., Graham, L., Grand-Pierre, N.,  
Hagos, B., Harford, A., Horton, L., Hulme, W., Iliev, I., Johnson, R.,  
Jones, C., Karatas, A., LaRocque, K., Lamazares, R., Landers, T.,  
Lehoczy, J., Levine, R., Liu, G., Maclean, C., MacDonald, P.,  
Marple, N., Matthews, C., McCarty, M., McEwan, P., McKernan, K.,  
McPherson, R., Meldrum, J., Menus, L., Mhova, T., Mlenga, V.,  
Murphy, T., Naylor, J., Nguyen, C., Norbu, C., Norman, C.H.,  
O'Connor, T., O'Donnell, P., O'Neill, D., Oliver, J., Peterson, K.,  
Phunkhang, P., Pierre, N., Pollara, V., Raymond, C., Retta, R.,  
Riedack, M., Riley, R., Rise, C., Rogov, P., Roman, J., Rosetti, M.,  
Roy, A., Santos, R., Schauer, S., Schnupack, R., Seaman, S., Severy, P.,  
Sounez, C., Spencer, B., Stange-Thomann, N., Stojanovic, N.,  
Strauss, N., Subramanian, A., Talamas, J., Testa, S., Theodore, J.,  
Travers, M., Travis, N., Trigg, W.J., Vassiliev, H., Viel, R., Vo, A.,  
Wilson, B., Wu, X., Wyman, D., Ye, W.J., Young, G., Zainoun, J.,  
Zemdek, L., Zimmer, A. and Zody, M.

Direct Submission  
Submitted (23-DEC-2000) Whitehead Institute/MIT Center for Genome  
Research, 320 Charles Street, Cambridge, MA 02141, USA  
3 (bases 1 to 100836)

Birren, B., Linton, L., Nusbaum, C., Lander, E., All, A., Allen, N.,  
Anderson, S., Barna, N., Baskien, V., Bloom, T., Boguslavsky, L.,  
Boukhgalter, B., Brown, A., Camata, J., Campolano, A., Chang, J.,  
Chazaro, B., Choepel, Y., Colangelo, M., Collins, S., Collimore, A.,  
Cook, A., Cooke, P., Dearellano, K., Dewar, K., Diaz, J.S., Dodge, S.,  
Faro, S., Ferreira, P., Fitzhugh, W., Gage, D., Galagan, J., Gardyna, S.,  
Ginde, S., Goid, S., Goyette, M., Graham, L., Grand-Pierre, N.,  
Hagos, B., Horton, L., Hulme, W., Iliev, I., Johnson, R., Jones, C.,  
Kamat, A., Karatas, A., Kelle, C., LaRocque, K., Lamazares, R.,  
Landers, T., Lehoczy, J., Levine, R., Lindblad-Toh, K., Liu, G.,  
Maclean, C., MacDonald, P., Major, J., Margulis, N., Matthews, L.,  
McCarthy, M., McEwan, P., McKernan, K., Meldrum, J., Menus, L.,  
Mhova, T., Mlenga, V., Murphy, T., Naylor, J., Nguyen, C., Nicol, R.,  
Norbu, C., Norman, C.H., O'Connor, T., O'Donnell, P., O'Neill, D.,  
Oliver, J., Peterson, K., Phunkhang, P., Pierre, N., Pollara, V.,  
Raymond, C., Retta, R., Riedack, M., Riley, R., Rise, C., Rogov, P.,  
Roman, J., Rosetti, M., Roy, A., Santos, R., Schauer, S., Schnupack, R.,  
Seaman, S., Severy, P., Spencer, B., Stange-Thomann, N., Stojanovic, N.,  
Strauss, N., Subramanian, A., Talamas, J., Testa, S., Theodore, J.,  
Topham, K., Travers, M., Travis, N., Trigg, W.J., Vassiliev, H.,  
Viel, R., Vo, A., Wilson, B., Wu, X., Wyman, D., Ye, W.J., Young, G.,  
Zainoun, J., Zemdek, L., Zimmer, A. and Zody, M.

Direct Submission  
Submitted (20-APR-2002) Whitehead Institute/MIT Center for Genome  
Research, 320 Charles Street, Cambridge, MA 02141, USA  
On Apr 20, 2002 this sequence version replaced gi:20177719.  
All repeats were identified using RepeatMasker:  
Smit, A.F.A. & Green, P. (1996-1997)  
<http://ftp.genome.washington.edu/RM/RepeatMasker.html>

Genome Center  
Center: Whitehead Institute/ MIT Center for Genome Research  
Center code: WIBR  
Web site: <http://www-seq.wi.mit.edu>  
Contact: [sequence.submissions@genome.wi.mit.edu](mailto:sequence.submissions@genome.wi.mit.edu)  
----- Project Information -----  
Center project name: LI2020  
Center clone name: 403\_G\_3  
----- Summary Statistics -----  
Sequencing vector: Plasmid; n/a; 100% of reads  
Chemistry: Dye-terminator Big Dye; 100% of reads  
Assembly program: Phrap; version 0.960731  
Consensus quality: 96277 bases at least Q40  
Consensus quality: 98051 bases at least Q30  
Consensus quality: 98838 bases at least Q20  
Insert size: 14000; agarose-fp  
Insert size: 99436; sum-of-consigs  
Quality coverage: 19.1 in Q20 bases; agarose-fp  
Quality coverage: 26.9 in Q20 bases; sum-of-consigs  
-----  
\* NOTE: This is a 'working draft' sequence. It currently  
\* consists of 15 contigs. Gaps between the contigs



\* are represented as runs of N. The order of the pieces  
 \* is believed to be correct as given, however the sizes  
 \* of the gaps between them are based on estimates that have  
 \* provided by the submittor.  
 \* This sequence will be replaced  
 \* by the finished sequence as soon as it is available and  
 \* the accession number will be preserved.

1 8098: contig of 8098 bp in length  
 8099 8198: gap of 100 bp  
 8199 8271: contig of 73 bp in length  
 8272 8371: gap of 100 bp  
 8372 9497: contig of 1126 bp in length  
 9498 9597: gap of 100 bp  
 9598 10863: contig of 1266 bp in length  
 10864 12232: contig of 1269 bp in length  
 12233 12332: gap of 100 bp  
 12333 13844: contig of 1512 bp in length  
 13845 13944: gap of 100 bp  
 13945 15215: contig of 1271 bp in length  
 15216 18234: contig of 2919 bp in length  
 18235 18334: gap of 100 bp  
 18335 21787: contig of 3453 bp in length  
 21788 21887: gap of 100 bp  
 21888 25485: contig of 3598 bp in length  
 25486 25585: gap of 100 bp  
 25586 40531: contig of 14946 bp in length  
 40532 40631: gap of 100 bp  
 40632 52119: contig of 11488 bp in length  
 52120 52219: gap of 100 bp  
 52220 64941: contig of 12722 bp in length  
 64942 65041: gap of 100 bp  
 65042 79587: contig of 14546 bp in length  
 79588 79687: gap of 100 bp  
 79688 100836: contig of 21149 bp in length.

## FEATURES

## source

1. .100836  
 /organism="Homo sapiens"  
 /mol\_type="genomic DNA"  
 /db\_xref="taxon:9606"  
 /chromosome="17"  
 /map="17"  
 /clone="RP11-403G3"  
 /clone\_1bp="RPC1-11 Human Male BAC"  
 1. .8098  
 /note="assembly\_fragment"  
 /note="end:SP6  
 vector\_side:left"  
 8199. .8271  
 /note="assembly\_fragment"  
 8372. .9497  
 /note="assembly\_fragment"  
 9598. .10863  
 /note="assembly\_fragment"  
 10964. .12232  
 /note="assembly\_fragment"  
 12333. .13844  
 /note="assembly\_fragment"  
 13945. .15215  
 /note="assembly\_fragment"  
 15316. .18234  
 /note="assembly\_fragment"  
 18335. .21787  
 /note="assembly\_fragment"  
 21888. .25485  
 /note="assembly\_fragment"  
 25586. .40531  
 /note="assembly\_fragment"  
 40632. .52119  
 /note="assembly\_fragment"  
 52220. .64941  
 /note="assembly\_fragment"

misc\_feature 65042..79587  
 /note="assembly\_fragment"  
 misc\_feature 79688..100836  
 /note="assembly\_fragment"  
 BASE COUNT 23052 a 27082 c 27053 g 22249 t 1400 others  
 ORIGIN

Query Match 99.7%; Score 1076.8; DB 2; Length 100836;  
 Best Local Similarity 99.8%; Pred. No. 2,7e-169;  
 Matches 1078; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 1 ATGGCCGAGCCATTCTTGTCAAGATTTCACACCCAGCTCAAACTGCACTGTACAGGT 60  
 DB 70785 ATGGCCGAGCCATTCTTGTCAAGATTTCACACCCAGCTCAAACTGCACTGTACAGGT 70726  
 QY 61 GCTGCTCTGTCCAGAGAGCTGAACCTTGAGCCGCCCGGAGCGCGAGAGCGGGT 120  
 DB 70725 GCTGCTCTGTCCAGAGAGCTGAACCTTGAGCCGCCCGGAGCGCGAGAGCGGGT 70666  
 QY 121 CCCGAGAGAGACGTAGTGGCAATCGAGAGCGTTCCCGAGTTGGTGGCCCTCCGGGG 180  
 DB 70665 CCCGAGAGAGACGTAGTGGCAATCGAGAGCGTTCCCGAGTTGGTGGCCCTCCGGGG 70606  
 QY 181 CCGAGAGGGGAGAGAGCTGGAATCCCAACCTCTCTTGCAGACCCAGGCTGTCCA 240  
 DB 70605 CCGAGAGGGGAGAGAGCTGGAATCCCAACCTCTCTTGCAGACCCAGGCTGTCCA 70546  
 QY 241 GAATCAGCTGCTGAGAGAGGGCGAGAGGGCGAATGGGGAGCACTGTCGGCTGCG 300  
 DB 70545 GAATCAGCTGCTGAGAGAGGGCGAGAGGGCGAATGGGGAGCACTGTCGGCTGCG 70486  
 QY 301 GGGCAGCTTCCGCGCGCGCAGAGTGAACCGACGCCGAGGCGCAGTGTGCTGCCAG 360  
 DB 70485 GGGCAGCTTCCGCGCGCGCAGAGTGAACCGACGCCGAGGCGCAGTGTGCTGCCAG 70426  
 QY 361 CCTTGTCAATGACTCCGAGGCCAGTAACTTGGGGGCTCTGCGGCGAGGGGCGAAGAGAG 420  
 DB 70425 CCTTGTCAATGACTCCGAGGCCAGTAACTTGGGGGCTCTGCGGCGAGGGGCGAAGAGAG 70366  
 QY 421 TGGGGACAGCAGAGAGACAGCTGGGGAGAGAAAAAATAGAGACGCCGTCAGAGAG 480  
 DB 70365 TGGGGACAGCAGAGAGACAGCTGGGGAGAGAAAAAATAGAGACGCCGTCAGAGAG 70306  
 QY 481 AAGCGCATTGGAAACCGTACTACAGCTGAACCTGGAGAGAGAGAAAAAGTTCCAGCAG 540  
 DB 70305 AAGCGCATTGGAAACCGTACTACAGCTGAACCTGGAGAGAGAGAAAAAGTTCCAGCAG 70246  
 QY 541 AAACAGAGCTTTCAGCTTCAAGATCCAGCCGAGATTTGGCCCAAGGGCCAGCCGCTC 600  
 DB 70245 AAACAGAGCTTTCAGCTTCAAGATCCAGCCGAGATTTGGCCCAAGGGCCAGCCGCTC 70186  
 QY 601 GGGCCCTATACACACGAGTTTCTCATGATGATACGACACGAGAGAGCGGATCTC 660  
 DB 70185 GGGCCCTATACACACGAGTTTCTCATGATGATACGACACGAGAGAGCGGATCTC 70126  
 QY 661 AAAACCGGCTGTACTCAAGCGGGCGCGCCCAATCCGACGACACCGCATGACGAC 720  
 DB 70125 AAAACCGGCTGTACTCAAGCGGGCGCGCCCAATCCGACGACACCGCATGACGAC 70066  
 QY 721 TTCATGAGAAAGGGGGTGAAGAGATGGGGCAGCGATGGATGGAGAGGAGCGCAC 780  
 DB 70065 TTCATGAGAAAGGGGGTGAAGAGATGGGGCAGCGATGGATGGAGAGGAGCGCAC 70006  
 QY 781 GAGTTTTCAGCAGCGGACTTCTCGAGAGCTACGAGCGGTACCAACCGAGAGCGTCCAG 840  
 DB 70005 GAGTTTTCAGCAGCGGACTTCTCGAGAGCTACGAGCGGTACCAACCGAGAGCGTCCAG 69946  
 QY 841 AACATAGAGAGAGAGCTCATCAAGAGTACTCTGGAACCTGGAAGAGTCTCTCGCGC 900  
 DB 69945 AACATAGAGAGAGAGCTCATCAAGAGTACTCTGGAACCTGGAAGAGTCTCTCGCGC 69886  
 QY 901 ATGAGAGAGAGAGAGAGCGGCTGCGGCTGAGAGAGAGCGGCTGAGTGGCGACGACGCG 960



```

Db      69885 ATGAGAGACGAGAACAAACGGCTGGCGCTGGAGAGCAACGGCTGGGTCGACGACGCG 69826
Qy      961 CGTGTCCGGAGAGCTGAGAGCTGAGAGCTGAGAGCTGGCGCGCCGAGAACCTCCAGCTGCTG 1020
Db      69825 CGTGTCCGGAGAGCTGAGAGCTGAGAGCTGAGAGCTGGCGCGCCGAGAACCTCCAGCTGCTG 69766
Qy      1021 ACCGAGAACGAACTGCAACCGGACAGAGAGCGAGCGCGCTTTCCAAATTGGAGACTAG 1080
Db      69765 ACCGAGAACGAACTGCAACCGGACAGAGAGCGAGCGCGCTTTCCAAATTGGAGACTAG 69706

RESULT 6
LOCUS   BD160090          2178 bp      DNA          linear      PAT 17-JAN-2003
DEFINITION
Primer for synthesizing full-length cDNA and use thereof.
ACCESSION
BD160090.1 GI:27965848
KEYWORDS
JP 2002191363-A/14933.
SOURCE
Homo sapiens (human)
ORGANISM
Homo sapiens
Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Euteleostomi;
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
1 (bases 1 to 2178)
Oca,T., Isogai,T., Nishikawa,T., Hayashi,K., Saito,K., Yamamoto,J.,
Ishii,S., Sugiyama,T., Wakamatsu,A., Nagai,K. and Otsuki,T.
Primer for synthesizing full-length cDNA and use thereof
Patent: JP 2002191363-A 14933 09-JUL-2002;
HELIX RESEARCH INSTITUTE
OS      Homo sapiens (human)
PN      JP 2002191363-A/14933
PD      09-JUL-2002
PF      28-JUL-2000 JP 2002280990
PI      TOSHIO OTA,TAKAO ISOGAI,TETSUO NISHIKAWA,KOJI HAYASHI,KAOBU
PI      SAITO
PI      JUNICHI YAMAMOTO,SHIZUKO ISHII,TOMOYASU SUGIYAMA,AI WAKAMATSU,
PI      KEIICHI NAGAI,TETSUJI OTSUKI
PC      C12N15/09,C07K14/47,C07K16/18,C12N1/15,C12N1/19,C12N1/21,C12N5/ PC
10' PC C12P21/02,C12O1/68//C12P21/08,G06F17/30,C12N15/00,C12N5/00 CC
Primer for synthesizing full-length cDNA and use thereof FH Key
Location/Qualifiers
(716) . (1792) .
FT      CDS
source   1. 2178
          /organism="Homo sapiens"
          /mol_type="genomic DNA"
          /db_xref="taxon:9606"
BASE COUNT      525 a      513 c      678 g      462 t
ORIGIN
Query Match      99.6%; Score 1075.2; DB 6; Length 2178;
Best Local Similarity 99.7%; Pred. No. 8,4e-169;
Matches 1077; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

Qy      1 ATGGCCGAGCCATTCTTGTGAGATATCAACACAGCCCTGAACTAGCACTGATACAGGT 60
Db      716 ATGGCCGAGCCATTCTTGTGAGATATCAACACAGCCCTGAACTAGCACTGATACAGGT 775
Qy      61 GCTGCTGCTGTCCAGGAAGAGCTGAACCTGAGCGCCCCCAGCGCGGAGAGCGGGT 120
Db      776 GCTGCTGCTGTCCAGGAAGAGCTGAACCTGAGCGCCCCCAGCGCGGAGAGCGGGT 835
Qy      121 CCGGAGAGAGACAGTAGTGTGGCAATCGAGAGCTTCCCCCAAGTTGGTGGCTCGGGG 180
Db      836 CCGGAGAGAGACAGTAGTGTGGCAATCGAGAGCTTCCCCCAAGTTGGTGGCTCGGGG 895
Qy      181 CCGGAGAGAGAGAGAGAGCTGGAATCCCAACCACTCTTGGAGACCGAGGCGTGTCCA 240
Db      896 CCGGAGAGAGAGAGAGAGCTGGAATCCCAACCACTCTTGGAGACCGAGGCGTGTCCA 955
Qy      241 GAATCTAGCTGCTGAGAGAGGGGAGAGAGGGCCAGATGGGAGCGAATTGTCGCTGGC 300

```

```

Db      956 GAATCTAGCTGCTGAGAGAGGGCCAGAGAGGCCAAGATGGGAGCGACTGTCCTCGCTGCG 1015
Qy      301 GCGGACTTCCCGCCCGCCAGAGAGTGAACCCAGCGCCGAGCGGAGCTGTGCTGCCAG 360
Db      1016 GCGGACTTCCCGCCCGCCAGAGAGTGAACCCAGCGCCGAGCGGAGCTGTGCTGCCAG 1075
Qy      361 CTTGTGATGACTCCGAGGCGAGTAAGTTGGGGGCTCTGCGCGAGGGGCGAAGAGAG 420
Db      1076 CTTGTGATGACTCCGAGGCGAGTAAGTTGGGGGCTCTGCGCGAGGGGCGAAGAGAG 1135
Qy      421 TGGGAGACGACAGACAGACAGCTGGGGAGAAAAAAGTAAGACGCCCGTCCAAAGAG 480
Db      1136 TGGGAGACGACAGACAGACAGCTGGGGAGAAAAAAGTAAGACGCCCGTCCAAAGAG 1195
Qy      481 AAGCGGCAATTGGAAAAACCGTACTACAAAGCTGAAGCTGGGAGAGAAAAAGTTTCAGAG 540
Db      1196 AAGCGGCAATTGGAAAAACCGTACTACAAAGCTGAAGCTGGGAGAGAAAAAGTTTCAGAG 1255
Qy      541 AAACAGAGCTTTCGAGCTTCAAGATCCGAGCGCGAGATTTGCCAAGGCGCAAGCGGTC 600
Db      1256 AAACAGAGCTTTCGAGCTTCAAGATCCGAGCGCGAGATTTGCCAAGGCGCAAGCGGTC 1315
Qy      601 GGGCCCTATTAACACACAGCGAGTCTCAGTGAATGATCAAGACAGAGAGCGGATCTC 660
Db      1316 GGGCCCTATTAACACACAGCGAGTCTCAGTGAATGATCAAGACAGAGAGCGGATCTC 1375
Qy      661 AAACCGGCTGTACTTCCAAAGCGGCGCGCCCAATCCGAGACACACAGCGATGACGAC 720
Db      1376 AAACCGGCTGTACTTCCAAAGCGGCGCGCCCAATCCGAGACACACAGCGATGACGAC 1435
Qy      721 TTCATGAGAAAGAGGGGTGAGAGAGATGGGGCAGCGATGAGATGGAGAGGAGCGGACG 780
Db      1436 TTCATGAGAAAGAGGGGTGAGAGAGATGGGGCAGCGATGAGATGGAGAGGAGCGGACG 1495
Qy      781 GAGTTTCTGACACGGGAGCTTCTCGAGAGCTACAGCGGTACCACACAGAGAGCTGCGAG 840
Db      1496 GAGTTTCTGACACGGGAGCTTCTCGAGAGCTACAGCGGTACCACACAGAGAGCTGCGAG 1555
Qy      841 AACATGAGACAGACAGAGCTCATCAAGAGTACTGGAAGCTGGAAGAGTGGCTCTGCGCG 900
Db      1556 AACATGAGACAGACAGAGCTCATCAAGAGTACTGGAAGCTGGAAGAGTGGCTCTGCGCG 1615
Qy      901 ATGAGAGACGAGAACAAACCGGCTGCGGCTGAGAGCAACCGGCTGGTGGCGACAGCGG 960
Db      1616 ATGAGAGACGAGAACAAACCGGCTGCGGCTGAGAGCAACCGGCTGGTGGCGACAGCGG 1675
Qy      961 CGTGTCCGGAGAGCTGAGAGCTGAGAGCTGAGAGCTGAGAGCTGAGAGCTGAGAGCTG 1020
Db      1676 CGTGTCCGGAGAGCTGAGAGCTGAGAGCTGAGAGCTGAGAGCTGAGAGCTGAGAGCTG 1735
Qy      1021 ACCGAGAACGAACTGCAACCGGACAGAGAGCGAGCGCGCTTCCAAATTGGAGACTAG 1080
Db      1736 ACCGAGAACGAACTGCAACCGGACAGAGAGCGAGCGCGCTTCCAAATTGGAGACTAG 1795

RESULT 7
LOCUS   AK023624          2178 bp      mRNA          linear      PRI 01-AUG-2002
DEFINITION
Homo sapiens cDNA FLJ13562 fis, clone PLAC81008080, highly similar
to Homo sapiens mRNA for HEXIM1 protein.
ACCESSION
AK023624
VERSION
AK023624.1 GI:10435606
KEYWORDS
oligo capping; fis (full insert sequence).
SOURCE
Homo sapiens (human)
ORGANISM
Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
1 (bases 1 to 2178)
Oca,T., Ota,T., Hayashi,K., Sugiyama,T., Otsuki,T., Suzuki,Y.,
Nishikawa,T., Nagai,K., Sugano,S., Shiratori,A., Sudo,H.,
Wakamatsu,M., Hosokita,T., Kaku,Y., Kodaira,H., Kondo,H.,
Sugawara,M., Takahashi,M., Chiba,Y., Ishida,S., Murakawa,K.,
Ono,Y., Takiguchi,S., Watanabe,S., Kimura,K., Murakami,K.,

```

Iehi, S., Kawai, Y., Saito, K., Yamamoto, J., Wakamatsu, A.,  
 Nakamura, Y., Nagahara, K., Masuno, Y., Ninomiya, K. and Iwayanagi, T.  
 NEDO human cDNA sequencing project  
 Unpublished  
 2 (bases 1 to 2178)  
 Isogai, T. and Otsuki, T.  
 Direct Submission  
 Submitted (23-AUG-2000) Takao Isogai, Helix Research Institute,  
 Genomics Laboratory, 1532-3 Yana, Kisarazu, Chiba 292-0812, Japan  
 (E-mail: genomics@nri.co.jp, Tel: 81-438-52-3975, Fax: 81-438-52-3986)  
 NEDO human cDNA sequencing project supported by Ministry of  
 International Trade and Industry of Japan, cDNA full insert  
 sequencing: Research Association for Biotechnology, cDNA library  
 construction, 5'- & 3'-end one pass sequencing and clone selection:  
 Helix Research Institute (supported by Japan Key Technology Center  
 etc.) and Department of Virology, Institute of Medical Science,  
 University of Tokyo.  
 Location/Qualifiers  
 1..2178  
 /organism="Homo sapiens"  
 /mol\_type="mRNA"  
 /db\_xref="taxon:9606"  
 /clone="PLACE1008080"  
 /tissue\_type="Placenta"  
 /clone\_id="PLACE1"  
 /note="cloning vector: PME18SFL3"  
 BASE COUNT 525 a 513 c 678 g 462 t  
 ORIGIN  
 Query Match 99.6%; Score 1075.2; DB 9; Length 2178;  
 Best Local Similarity 99.7%; Pred. No. 8,4e-169;  
 Matches 1077; Conservative 0; Mismatches 3; Indels 0; Gaps 0;  
 1 ATGGCCGAGCATTTCTTGCAGATATCAACACGAGCTCAAACTGACATGACAGGT 60  
 716 ATGGCCGAGCATTTCTTGCAGATATCAACACGAGCTCAAACTGACATGACAGGT 775  
 61 GCTGCTGCTGTCCAGAGAGAGTGAACCTGAGCGCCGCCAGCGCGAGAGAGCGGGT 120  
 776 GCTGCTGCTGTCCAGAGAGAGTGAACCTGAGCGCCGCCAGCGCGAGAGAGCGGGT 835  
 121 CCCGAGAGAGACATGATGGTGGCAATGAGAGCTTCCCGCATGGTGGTGGTCCGGGG 180  
 836 CCCGAGAGAGACATGATGGTGGCAATGAGAGCTTCCCGCATGGTGGTGGTCCGGGG 895  
 181 CCCGAGAGAGAGAGAGAGTGAATCCCAACCACTCCCTTGCAGACCCAGGCGCTGTCCA 240  
 896 CCCGAGAGAGAGAGAGAGTGAATCCCAACCACTCCCTTGCAGACCCAGGCGCTGTCCA 955  
 241 GAATCTAGCTGCTGAG 300  
 956 GAATCTAGCTGCTGAG 1015  
 301 GGGGAGTTCCCGCCCGCCGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 360  
 1016 GGGGAGTTCCCGCCCGCCGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1075  
 361 CCTTGCATGACTCCGAGGCGAGTAAAGTGGGGGCTCTGCGCGAGGGGGCGAAGAGAG 420  
 1076 CCTTGCATGACTCCGAGGCGAGTAAAGTGGGGGCTCTGCGCGAGGGGGCGAAGAGAG 1135  
 421 TGGGAGACAG 480  
 1136 TGGGAGACAG 1195  
 481 AAGCGCATTTGGAACCGTACTCAAGCTGAACTGGGAGAGAGAGAGAGAGAGAGAGAG 540  
 1196 AAGCGCATTTGGAACCGTACTCAAGCTGAACTGGGAGAGAGAGAGAGAGAGAGAGAG 1255  
 541 AAGCAGAGCTTGCAGCTTCAAGAGATCCGAGCGGAGATGTTGGCCAAAGGCGACCGCGTC 600  
 1256 AAGCAGAGCTTGCAGCTTCAAGAGATCCGAGCGGAGATGTTGGCCAAAGGCGACCGCGTC 1315

601 GGGCCCTATTAACACGAGAGTTCCTCATGATGATCAAGACAGAGAGAGAGAGAGAGATC 660  
 1316 GGGCCCTATTAACACGAGAGTTCCTCATGATGATCAAGACAGAGAGAGAGAGAGATC 1375  
 661 AAAACCGGCTGTACTCCAAAGGCGCGCGCCCAATCCGAGACAGACCCAGCATGACGAC 720  
 1376 AAAACCGGCTGTACTCCAAAGGCGCGCGCCCAATCCGAGACAGACCCAGCATGACGAC 1435  
 721 TTTCATGAG 780  
 1436 TTTCATGAG 1495  
 781 GAGTTTCTGACGCGGAGCTTCTCGAGACGTAACGCGGTATCAACAGAGAGAGAGAGAG 840  
 1496 GAGTTTCTGACGCGGAGCTTCTCGAGACGTAACGCGGTATCAACAGAGAGAGAGAGAG 1555  
 841 AACATGAGCAACAG 900  
 1556 AACATGAGCAACAG 1615  
 901 ATGAGAGACAG 960  
 1616 ATGAGAGACAG 1675  
 961 CCGTGGCGGAG 1020  
 1676 CCGTGGCGGAG 1735  
 1021 ACCGAG 1080  
 1736 ACCGAG 1795

RESULT 8  
 LOCUS BC006460 2086 bp mRNA linear PRI 12-JUL-2001  
 DEFINITION Homo sapiens, HMBA-inducible, clone MGC:1880 IMAGE:3535529, mRNA,  
 complete cds.  
 ACCESSION BC006460  
 VERSION BC006460.1 GI:13623668  
 KEYWORDS MGC.  
 SOURCE Homo sapiens (human)  
 ORGANISM Homo sapiens  
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 Mammalia; Eutheria; Primates; Catarrhini; Homiidae; Homo.  
 1 (bases 1 to 2086)  
 Struhsberg, R.  
 Direct Submission  
 Submitted (09-APR-2001) National Institutes of Health, Mammalian  
 Gene Collection (MGC), Cancer Genomics Office, National Cancer  
 Institute, 31 Center Drive, Room 11A03, Bethesda, MD 20892-2590,  
 USA  
 NIH-MGC Project URL: <http://mgc.nci.nih.gov>  
 CONTACT Contact: MGC help desk  
 Email: [cgapbs-remail.nih.gov](mailto:cgapbs-remail.nih.gov)  
 Tissue Procurement: DCTD/DRP  
 CDNA Library Preparation: Rubin Laboratory  
 DNA Sequencing by: Genome Sequence Centre,  
 BC Cancer Agency, Vancouver, BC, Canada  
 info@bcgsc.bc.ca  
 Steven Jones, Jennifer Asano, Ian Bosdet, Yaron Butterfield,  
 Susanna Chan, Readman Chiu, Chris Fjell, Erin Garland, Ran Guin,  
 Leticia Heiso, Martin Krzyzinski, Reta Kutsche, Oliver Lee, Soo  
 Sen Lee, Victor Ling, Carrie Mathewson, Candice McLeary, Steven  
 Ness, Pawan Pandoh, Anna-Lisa Prabhu, Parveen Seeded, Jacqueline  
 Schein, Diane Smalhus, Michael Smith, Lorraine Spence, Jeff Stolt,  
 Michael Thorne, Miranada Teal, Natasja van den Bosch, Jill Vardy,  
 George Yang, Scott Zuyderduyn, Marco Marra.

Clome distribution: MGC clone distribution information can be found  
 through the I.M.A.G.E. Consortium/LLNL at: <http://image.llnl.gov>  
 Series: IRAL Plate: 8 Row: d Column: 1

This clone was selected for full length sequencing because it passed the following selection criteria: matched mRNA gi: 4062855.

## FEATURES

source

1.2086  
/organism="Homo sapiens"  
/mol\_type="mRNA"  
/db\_xref="taxon:9606"  
/db\_xref="taxon:9606"  
/clone="MGC:1880 IMAGE:353529"  
/issue\_type="Lung, small cell carcinoma"  
/clone\_id="N1H MGC 7"  
/lab\_host="DH10B-R"  
/note="vector: pOT17"  
606..1685  
/codon\_start=1  
/product="HMB-Inducible"  
/protein\_id="AAH06460.1"  
/db\_xref="GI:13623669"  
/translation="MAEPFLSEYQHQPOTSNCTGAANAQVEELNPERPPGAERVPED  
SFWOSAPFQLOGRPGPBGESLSEPPLOQACBESSCLREGEKQMGDDSSAGD  
FPPAVEPTPAELIACPDHSEASKLGAAPAGEEEMQOQOLGKKHRRSKK  
KHHMXYKLTWEKKKKFKDEKSLRASRIAEAFKAGVAVYNTTQFLMDHDEEP  
DLKTGYKRAAKSDPTSDDDPMEGESEDSGDMGDSSEFLORDPSFYRYHT  
BSLQNSKQELIKETLEKCKSRMEDENNRLRLSKRLGDDARVLELELDRLRA  
ENLQLTLELHROQERAPLSKFGD"

## CDS

BASE COUNT 510 a 505 c 618 g 453 t  
ORIGIN

Query Match 99.4%; Score 1073.6; DB 9; Length 2086;  
Best Local Similarity 99.6%; Pred. No. 1.6e-168;  
Matches 1076; Conservative 0; Mismatches 4; Indels 0; Gaps 0;

1 ATGGCCGACCATCTTCTGACATATCAACACCACTCAAACTGACATGACAGT 60  
606 ATGGCCGACCATCTTCTGACATATCAACACCACTCAAACTGACATGACAGT 665  
61 GCTGCTGCTGTCAGAGAGAGTGAACCTTACGCGCCCGCAGCGCGAGAGCGGGTG 120  
666 GCTGCTGCTGTCAGAGAGAGTGAACCTTACGCGCCCGCAGCGCGAGAGCGGGTG 725  
121 CCCGAGAGAGAGAGTGAAGTGAAGTGAAGTGAAGTGAAGTGAAGTGAAGTGAAGT 180  
726 CCCGAGAGAGAGAGTGAAGTGAAGTGAAGTGAAGTGAAGTGAAGTGAAGTGAAGT 785  
181 CCCGAGAGAGAGAGAGTGAAGTGAAGTGAAGTGAAGTGAAGTGAAGTGAAGTGAAGT 240  
786 CCCGAGAGAGAGAGAGTGAAGTGAAGTGAAGTGAAGTGAAGTGAAGTGAAGTGAAGT 845  
241 GAATCTAGCTGCTGTCAGAGAGAGAGTGAAGTGAAGTGAAGTGAAGTGAAGTGAAGT 300  
846 GAATCTAGCTGCTGTCAGAGAGAGAGTGAAGTGAAGTGAAGTGAAGTGAAGTGAAGT 905  
301 GGGGACTTCCCGCGCGCGAGAGAGTGAAGTGAAGTGAAGTGAAGTGAAGTGAAGTGAAGT 360  
906 GGGGACTTCCCGCGCGCGAGAGAGTGAAGTGAAGTGAAGTGAAGTGAAGTGAAGTGAAGT 965  
361 CCTTTCATGACTCCGAGAGAGAGTGAAGTGAAGTGAAGTGAAGTGAAGTGAAGTGAAGT 420  
966 CCTTTCATGACTCCGAGAGAGAGTGAAGTGAAGTGAAGTGAAGTGAAGTGAAGTGAAGT 1025  
421 TGGGAGACAGAGAGAGAGTGAAGTGAAGTGAAGTGAAGTGAAGTGAAGTGAAGTGAAGT 480  
1026 TGGGAGACAGAGAGAGAGTGAAGTGAAGTGAAGTGAAGTGAAGTGAAGTGAAGTGAAGT 1085  
481 AAGCGCATTTGGAACCGTACCTGAACTGAACTGAACTGAACTGAACTGAACTGAACTGAACT 540  
1086 AAGCGCATTTGGAACCGTACCTGAACTGAACTGAACTGAACTGAACTGAACTGAACTGAACT 1145  
541 AAGCGCATTTGGAACCGTACCTGAACTGAACTGAACTGAACTGAACTGAACTGAACTGAACT 600  
1146 AAGCGCATTTGGAACCGTACCTGAACTGAACTGAACTGAACTGAACTGAACTGAACTGAACT 1205  
601 GCGCCCTATTAACAACAGCATTTCTCATGATGATCAAGACAGAGAGAGCGGATCTC 660

Db 1206 GCGCCCTATTAACAACAGCATTTCTCATGATGATCAAGACAGAGAGAGCGGATCTC 1265  
Qy 661 AAAACCGGCTGTACTCCAGAGAGAGCGCCCGCAATCCGACGACACAGAGATGACGAC 720  
Db 1266 AAAACCGGCTGTACTCCAGAGAGAGCGCCCGCAATCCGACGACACAGAGATGACGAC 1325  
Qy 721 TTGATGAG 780  
Db 1326 TTGATGAG 1385  
Qy 781 GAGTTTTCGACCGGAGCTTCTCGAGAGAGTACGAGCGGTACCAACGAGAGCGCTGAG 840  
Db 1386 GAGTTTTCGACCGGAGCTTCTCGAGAGAGTACGAGCGGTACCAACGAGAGCGCTGAG 1445  
Qy 841 AACATGACCAAGCAGAGAGCTCATCAAGAGTACTGGAATCTGAGAGAGAGAGAGAGAGAG 900  
Db 1446 AACATGACCAAGCAGAGAGCTCATCAAGAGTACTGGAATCTGAGAGAGAGAGAGAGAGAG 1505  
Qy 901 ATGAG 960  
Db 1506 ATGAG 1565  
Qy 961 CCGTGGCGGAGAGCTGAGAGTGAAGTGAAGTGAAGTGAAGTGAAGTGAAGTGAAGTGAAGT 1020  
Db 1566 CCGTGGCGGAGAGCTGAGAGTGAAGTGAAGTGAAGTGAAGTGAAGTGAAGTGAAGTGAAGT 1625  
Qy 1021 ACCGAG 1080  
Db 1626 ACCGAG 1685

RESULT 9  
AC024047/c  
LOCUS Homo sapiens chromosome 17 clone RP11-403G3, WORKING DRAFT  
DEFINITION  
ACCESSION AC024047  
VERSION AC024047.3 GI:8570008  
KEYWORDS HTG; HTGS PHASE1; HTGS\_DRAFT.  
SOURCE Homo sapiens (human)  
ORGANISM Homo sapiens  
REFERENCE  
AUTHORS Waterston, R.H.  
TITLE The sequence of Homo sapiens clone  
JOURNAL Unpublished  
REFERENCE  
AUTHORS Waterston, R.H.  
TITLE Direct Submission  
JOURNAL Submitted (20-FEB-2000) Genome Sequencing Center, Washington University School of Medicine, 444 Forest Park Parkway, St. Louis, MO 63108, USA  
COMMENT  
On Jun 16, 2000 this sequence version replaced gi:7109599.

----- Genome Center -----  
Center: Washington University Genome Sequencing Center  
Center code: WUGSC  
Web site: http://genome.wustl.edu/gsc/index.shtml  
----- Project Information -----  
Center project name: H NH040303  
----- Summary Statistics -----  
Sequencing vector: M13, 100%  
Sequencing method: Plasmid, 0%  
Chemistry: Dye-terminator Big Dye, 0% of reads  
Assembly program: Phrap, version 0.990319  
Consensus quality: 127942 bases at least Q40  
Consensus quality: 133108 bases at least Q30  
Consensus quality: 135734 bases at least Q20  
Insert size: 147000; agarose-IP  
Insert size: 139626; sum-of-contrigs

Quality coverage: 4.06 in Q20 bases; agarose-fp  
Quality coverage: 4.33 in Q20 bases; sum-of-contigs

NOTE: This is a 'working draft' sequence. It currently consists of 28 contigs. The true order of the pieces is not known and their order in this sequence record is arbitrary. Gaps between the contigs are represented as runs of N, but the exact sizes of the gaps are unknown. This record will be updated with the finished sequence as soon as it is available and the accession number will then be preserved.

1	1575:	contig of 1575 bp in length
1576	1675:	gap of unknown length
1576	3114:	contig of 1439 bp in length
3115	3214:	gap of unknown length
3215	5389:	contig of 2175 bp in length
5390	5489:	gap of unknown length
5490	7159:	contig of 1640 bp in length
7130	7229:	gap of unknown length
7230	9240:	contig of 2011 bp in length
9241	9340:	gap of unknown length
9341	11177:	contig of 1837 bp in length
11178	11277:	gap of unknown length
11278	13113:	contig of 1836 bp in length
13114	13213:	gap of unknown length
13214	16526:	contig of 3313 bp in length
16527	16826:	gap of unknown length
16827	19864:	contig of 3238 bp in length
19865	19964:	gap of unknown length
19965	22667:	contig of 2703 bp in length
22668	22767:	gap of unknown length
22768	26650:	contig of 1883 bp in length
26651	26780:	gap of unknown length
26781	29303:	contig of 2553 bp in length
29304	29403:	gap of unknown length
29404	32073:	contig of 2670 bp in length
32074	32173:	gap of unknown length
32174	37141:	contig of 4968 bp in length
37142	37241:	gap of unknown length
37242	40718:	contig of 3497 bp in length
40739	40838:	gap of unknown length
40839	45166:	contig of 4328 bp in length
45167	45266:	gap of unknown length
45267	49501:	contig of 4235 bp in length
49502	49601:	gap of unknown length
49602	55647:	contig of 6046 bp in length
55648	55747:	gap of unknown length
55748	61532:	contig of 5785 bp in length
61533	61632:	gap of unknown length
61633	68005:	contig of 6373 bp in length
68006	68105:	gap of unknown length
68106	75211:	contig of 7106 bp in length
75212	75311:	gap of unknown length
75312	81560:	contig of 6249 bp in length
81561	81650:	gap of unknown length
81661	88524:	contig of 6864 bp in length
88625	88624:	gap of unknown length
88625	95519:	contig of 6895 bp in length
95520	95619:	gap of unknown length
95620	105148:	contig of 9529 bp in length
105249	115248:	gap of unknown length
115250	115369:	contig of 10021 bp in length
115370	128646:	gap of unknown length
128647	128746:	contig of 13277 bp in length
128747	1428346:	gap of unknown length
128747	1428346:	contig of 13580 bp in length

## FEATURES

```
source
1. 142326
/organism="Homo sapiens"
/mol_type="genomic DNA"
/db_xref="taxon:9606"
/chromosome="17"
/clone="RP11-403G3"
```

[illegible]

Query Match	97.5%;	Score 1052.8;	DB 2;	Length 142326;
Best Local Similarity	99.6%;	Pred. No. 2.5e-165;		
Matches 1076;	Conservative 0;	Mismatches 2;	Indels 2;	Gaps 2;

Qy	1	ATGGCCGACCAATTTCTTGTCAAAATATCAACACCGCTCTCAATATCAACTGTACAGGT	60
Db	108505	ATGGCCGACCAATTTCTTGTCAAAATATCAACACCGCTCTCAATATCAACTGTACAGGT	108446
Qy	61	GCTGCTGCTGTCCAGGAAGCTGAACCTTAGAGGCCCTCCAGGCGCGAGAGACGGGTG	120
Db	108445	GCTGCTGCTGTCCAGGAAGCTGAACCTTAGAGGCCCTCCAGGCGCGAGAGACGGGTG	108386
Qy	121	CCCGATGAGAGACAGTAGGTGGCAATCGAGAGCGTTCCTCCAGTTTGGGTGGACCGTCCGGG	180
Db	108385	CCCGATGAGAGACAGTAGGTGGCAATCGAGAGCGTTCCTCCAGTTTGGGTGGACCGTCCGGG	108326
Qy	181	CCGAGAGGGGAAAGGAGCCTTGAAATCCCAACCACTCCCTTTGCAAGCTCCAGGCTGTCCA	240
Db	108325	CCGAGAGGGGAAAGGAGCCTTGAAATCCCAACCACTCCCTTTGCAAGCTCCAGGCTGTCCA	108266
Qy	241	GAATCTAGCTGCTGAGAGAGGGCCGAGAAATGGGGACGAACTGTCCGCTGGC	300
Db	108265	GAATCTAGCTGCTGAGAGAGGGCCGAGAAATGGGGACGAACTGTCCGCTGGC	108206
Qy	301	GGCGAATTCCCGCCGCGCAGAAAGTGAACCGACGCCGAGCGCGAAGCTTCCGCCAG	360
Db	108205	-GGCAATTCCCGC-CGCGCGAGAAAGTGAACCGACGCCGAGCGCGAAGCTTCCGCCAG	108148
Qy	361	CCTTTCATAGACTCCGAGGGCCAGTAAAGTTGGGGGTCTCTGCGGAGGGGGCGAAAGAG	420
Db	108147	CCTTTCATAGACTCCGAGGGCCAGTAAAGTTGGGGGTCTCTGCGGAGGGGGCGAAAGAG	108088
Qy	421	TGGGGAACAACAGACAGACAGCTGGGGAAGAAAAACAATAAGACACCCGTCGAAGAG	480
Db	108087	TGGGGAACAACAGACAGACAGCTGGGGAAGAAAAACAATAAGAGACCCCGTCCAAAGAG	108028
Qy	481	AAGCGGCATTGGAACCCGTAACAGCTGAATCGGGAAGAGAAAGAAAGTTGCAGAG	540
Db	108027	AAGCGGCATTGGAACCCGTAACAGCTGAATCGGGAAGAGAAAGAAAGTTGCAGAG	107968
Qy	541	AAACAAGAGCTTCGAGCTTCAAGGATTCGAGCCGAGAGTTTCGCAAGGGCCAGCGGTC	600
Db	107967	AAACAAGAGCTTCGAGCTTCAAGGATTCGAGCCGAGAGTTTCGCAAGGGCCAGCGGTC	107908
Qy	601	GGCGCCCTAATACACACAGCAGATTCTCTCATGTGATATCAAGCCAGAGAGACCGGATCTC	660
Db	107907	GGCGCCCTAATACACACAGCAGATTCTCTCATGTGATATCAAGCCAGAGAGACCGGATCTC	107848
Qy	661	AAAAACGGGCTGTACTCCAAAGCGGGCGCGCCCAATTCGACGACACCAAGCATGAGAC	720
Db	107847	AAAAACGGGCTGTACTCCAAAGCGGGCGCGCCCAATTCGACGACACCAAGCATGAGAC	107788
Qy	721	TTTCATGAGAGAGGGGGGTGAGAGGATGAGGGGCAAGCTGGAGTGGAGGGGAGCGGAGC	780
Db	107787	TTTCATGAGAGAGGGGGGTGAGAGGATGAGGGGCAAGCTGGAGTGGAGGGGAGCGGAGC	107728
Qy	781	GAGTTTCTGACAGCGGAGCTTCTCGGAGACGTACGAGCGGTACCAACAGAGAGCTTCAG	840
Db	107727	GAGTTTCTGACAGCGGAGCTTCTCGGAGACGTACGAGCGGTACCAACAGAGAGCTTCAG	107668
Qy	841	AACATGAGAACACAGAGAGCTCAAAAGAGTAACTGGAACTGGAGAAAGTGCCTTCGCGC	900
Db	107667	AACATGAGAACACAGAGAGCTCAAAAGAGTAACTGGAACTGGAGAAAGTGCCTTCGCGC	107608
Qy	901	ATGAGAGACGAGAACCAACCGGCTGCGCTGGAGAGCAAGCGGCTGGGTGGCGAGAGCGG	960
Db	107607	ATGAGAGACGAGAACCAACCGGCTGCGCTGGAGAGCAAGCGGCTGGGTGGCGAGAGCGG	107548
Qy	961	CGTGTGCGGAGCTGGAGCTGGAGCTGGACCGGCTGCGCGCTCGAGAACTTTCAGCTTCTG	1020

Db 107547 CATTGCGGAGCTGAGCTGAGCTGAGCCGCGCTGCGCGAGAACTCCAGCTGCTG 107488  
QY 1021 ACCGAGAACGAACTGACCGCGAGCAGAGCGAGCGCGCTTCCAACTTTGGAGACTAG 1080  
Db 107487 ACCGAGAACGAACTGACCGCGAGCAGAGCGAGCGCGCTTCCAACTTTGGAGACTAG 107428

RESULT 10  
AY090614 3402 bp mRNA linear ROD 18-JUN-2002  
LOCUS Mus musculus cardiac lineage protein 1 (Clp1) mRNA, complete cds.  
DEFINITION  
ACCESSION AY090614  
VERSION AY090614.1 GI:20135640  
KEYWORDS  
SOURCE Mus musculus (house mouse)  
ORGANISM  
REFERENCE  
AUTHORS Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.  
TITLE Huang, F., Wagner, M., and Siddiqui, M.A.Q.  
CLP-1 gene  
Structure, expression, and functional characterization of the mouse  
CLP-1 gene  
JOURNAL Gene 292 (1-2), 245-259 (2002)  
MEDLINE 22114986  
PUBMED 12119119  
REFERENCE 2 (bases 1 to 3402)  
AUTHORS Huang, F., Wagner, M., and Siddiqui, M.  
TITLE Direct Submission  
JOURNAL Submitted (19-MAR-2002) Anatomy and Cell Biology, State University  
of New York Health Science Center, 450 Clarkson Avenue, Brooklyn,  
NY 11203, USA

FEATURES  
source  
1. .3402  
/organism="Mus musculus"  
/mol\_type="mRNA"  
/strain="129/SvJ"  
/db\_xref="taxon:10090"  
1. .3402  
/gene="Clp1"  
596. 1666  
/gene="Clp1"  
/note="nuclear protein; CLP-1"  
/codon\_start=1  
/product="cardiac lineage protein 1"  
/protein\_id="AA09026.1"  
/db\_xref="GI:20135641"  
translation="MAAPLLEHQPQTSNCTGAIVHEHTSERPPSAEERVPKED  
SRWQSRASLSGSGRPGQEGEGIKQLPPLQTNACPELSLEKGEKGNGEBDLSGGA  
SPEAGEPMSSESLVQPHSEATKQBAAPAGBEPGQOQRLGKKHRRRPSKKRH  
WKPYVLTWEKKKPKDEKOSLRASRVAFEMFAKGVAPYNTQFLMDHDOEPDLK  
TGLYPRRAAKSDTSDDEPVEAGREDDSGMGDGSFLOPDFSEYERYHAESTL  
QNSKQBLIKEYIELEKCLSRKEDENNRLHLESKRUGVDARVRELELELRALNTL  
QLTLENLEHQERRAPLSKFGD"

BASE COUNT 920 a 771 c 890 g 821 t  
ORIGIN

Query Match 77.9%; Score 840.8; DB 10; Length 3402;  
Best Local Similarity 86.9%; Pred. No. 6.1e-130;  
Matches 939; Conservative 0; Mismatches 132; Indels 9; Gaps 1;

QY 1 ATGGCCGAGCCATTCTTGCAGATATCAACACAGCCCTCAAACTGCACTGTACAGT 60  
Db 596 ATGGCCGAGCCATTCTTGCAGATATCAACACAGCCCTCAAACTGCACTGTACAGT 655  
QY 61 GTTGTCTGTCTCAAGAGAGCTGAACCTGAGCGCCCGCCAGCGCGAGAGCGGGT 120  
Db 656 GTTGTCTGTCTCAAGAGAGCTGAACCTGAGCGCCCGCCAGCGCGAGAGCGGGT 715  
QY 121 CCGGAGAGAGCACTAGTGGCAATCGAGACGCTTCCCAAGTTGTGGCGCTCCGGG 180  
Db 716 CCGGAGAGAGCACTAGTGGCAATCGAGACGCTTCCCAAGTTGTGGCGCTCCGGG 775  
QY 181 CCGGAGAGAGAGCACTAGTGGCAATCGAGACGCTTCCCAAGTTGTGGCGCTCCGGG 240

Db 776 CAGAGAGGAGAGG393GCTTGAAGACCAAGCTGCCCCCACTTGCAACAGATCTGTCCA 835  
QY 241 GAATCTAGCTGCTGAGAGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 300  
Db 836 GAATGAGCTCCCTGGAGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGG 895  
QY 301 GCGCACTTCCGCGCGCGCGAGAGTGAACCGAGCGCGAGAGGCGAGCTGCTGCCAG 360  
Db 896 GGTGCTCCCGCTCGCGCGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGG 946  
QY 361 CTTGTCTAGTCTCCGAGGCGAGTAACTTGGGGGCTCTCTGCGCGAGGGGCGAAGAGAG 420  
Db 947 CAGGTCTAGTCTCCGAGGCGAGTAACTTGGGGGCTCTCTGCGCGAGGGGCGAAGAGCA 1006  
QY 421 TGGGGACAGCAGCAGAGCAGCTGGGGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGG 480  
Db 1007 TGGGGACAGCAGCAGAGCAGCTGGGGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGG 1066  
QY 481 AAGCGCATTTGGAACCGTACTTACAGCTGAACCTGGGAGAGAGAGAGAGAGAGAGAG 540  
Db 1067 AAGCGCATTTGGAACCGTACTTACAGCTGAACCTGGGAGAGAGAGAGAGAGAGAGAG 1126  
QY 541 AAGCAGAGCTTTCAGACTTCAAGATCCGAGCGCGAGATTTGCCAAGGGCGAGCGGTC 600  
Db 1127 AAGCAGAGCTTTCAGACTTTCAGAGCTTTCAGAGCTTTCAGAGCTTTCAGAGCTT 1186  
QY 601 GGGCCCTTAAACACACAGAGTTCCTCATGATGATCAACAGAGAGAGGCGGATCTC 660  
Db 1187 GGGCCCTTAAACACACAGAGTTCCTCATGATGATCAACAGAGAGAGGCGGATCTC 1246  
QY 661 AAAACCGGCTGTACTTCCAAAGCGGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGAG 720  
Db 1247 AAAACCGGCTGTACTTCCAAAGCGGCGCGCGCGCGCGCGCGCGCGCGCGCGCGAG 1306  
QY 721 TTCAATGAGAGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGG 780  
Db 1307 TTGTGAGAGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGG 1366  
QY 781 GAGTTTTCGACGCGGAGCTTTCGAGAGCTGCGAGCGGTACCAACGAGAGCGCTGCAG 840  
Db 1367 GAGTTTTCGACGCGGAGCTTTCGAGAGCTGCGAGCGGTACCAACGAGAGCGCTGCAG 1426  
QY 841 AACATGAG 900  
Db 1427 AACATGAG 1486  
QY 901 ATGAG 960  
Db 1487 AAG 1546  
QY 961 CCGTTCGCGGAG 1020  
Db 1547 CCGTTCGCGGAG 1606  
QY 1021 ACCGAG 1080  
Db 1607 ACCGAG 1666

RESULT 11  
AL731805 111002 bp DNA linear ROD 21-JUN-2002  
LOCUS Mouse DNA sequence from clone RP23-358E19 on chromosome 11,  
DEFINITION  
ACCESSION AL731805  
VERSION AL731805.8 GI:21540125  
KEYWORDS  
SOURCE HTG.  
ORGANISM Mus musculus (house mouse)  
REFERENCE  
1 (bases 1 to 111002)  
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.

## AUTHORS

Dunn, M.  
Direct Submission  
Submitted (21-JUN-2002) Wellcome Trust Sanger Institute, Hinxton,  
Cambridgeshire, CB10 1SA, UK. E-mail enquiries:  
humquerry@sanger.ac.uk Clone requests: clonerequests@sanger.ac.uk  
On Jun 23, 2002 this sequence version replaced gi:21531483.

## COMMENT

----- Genome Center

Center: Wellcome Trust Sanger Institute  
Center code: SC  
Web site: <http://www.sanger.ac.uk>  
Contact: humquerry@sanger.ac.uk

During sequence assembly data is compared from overlapping clones. Where differences are found these are annotated as variations together with a note of the overlapping clone name. Note that the variation annotation may not be found in the sequence submission corresponding to the overlapping clone, as we submit sequences with only a small overlap as described above.

This sequence was finished as follows unless otherwise noted: all regions were either double-stranded or sequenced with an alternate chemistry or covered by high quality data (i.e., phred quality >= 30); an attempt was made to resolve all sequencing problems, such as compressions and repeats; all regions were covered by at least one plasmid subclone or more than one M13 subclone; and the assembly was confirmed by restriction digest. The following abbreviations are used to associate primary accession numbers given in the feature table with their source databases: Em, EMBL; SW, SWISSPROT; Tr, TREMBL; Wp, WORMPEP; Information on the WORMPEP database can be found at [http://www.sanger.ac.uk/Projects/C\\_elegans/wormpep](http://www.sanger.ac.uk/Projects/C_elegans/wormpep) RP23-35BE19 is from the RPEC-23 Mouse PAC Library constructed by the group of Pierre de Jong. For further details see <http://www.chori.org/bacpac/home.htm> VECTOR: pBACE3.6.

## FEATURES

## source

1. 111002  
/organism="Mus musculus"  
/mol\_type="genomic DNA"  
/db\_xref="taxon:10090"  
/chromosome="11"  
/clone="RP23-35BE19"  
/clone\_1kb="RPEC-23"

BASE COUNT 28051 a 27579 c 27135 g 28237 t

## ORIGIN

Query Match 77.9%; Score 840.8; DB 10; Length 111002;  
Best Local Similarity 86.9%; Pred. No. 3.8e-130;  
Matches 939; Conservative 0; Mismatches 132; Indels 9; Gaps 1;

QY 1 ATGGCCGAGCCATTCTTGTGAGATATCAACACAGCCCTCAAACTAGCAACTGTACAGGT 60  
DB 91440 ATGGCCGAGCCATTCTTGTGAGATATCAACACAGCCCTCAAACTAGCAACTGTACAGGT 91499  
QY 61 GCTGCTGCTCTCCAGAGAGAGCTGAACCTGAGGCGCCCGAGCGCGGAGAGCGGGTG 120  
DB 91500 GCTGCTGCTCTCCAGAGAGAGCTGAACCTGAGGCGCCCGAGCGCGGAGAGCGGGTG 91559  
QY 121 CCGGAG 180  
DB 91560 CCGGAG 91619  
QY 181 CCGGAG 240  
DB 91620 CAGGAG 91679  
QY 241 GAATCTTACTGCTCTGAG 300  
DB 91680 GAATCTTACTGCTCTGAG 91739  
QY 301 GGGAG 360  
DB 91740 GGGAG 91790

QY 361 CTTGTCATGACTCCGAGGCGAGTAAGTTGGGGCTCTTGGCCGAGGGCCGAGAGAGAG 420  
DB 91791 CCGAGTCATGACTCCGAGGCGAGTAAGTTGGGGCTCTTGGCCGAGGGCCGAGAGAGAG 91850  
QY 421 TGGGAGACAG 480  
DB 91851 TGGGAGACAG 91910  
QY 481 AAGCGCATTTGGAACACCGTACTACAACTGAACTGGAGAGAGAGAGAGAGAGAGAGAG 540  
DB 91911 AAGCGCATTTGGAACACCGTACTACAACTGAACTGGAGAGAGAGAGAGAGAGAGAG 91970  
QY 541 AAACAGAGCTTGTGAGACTTCAAGATCCGAGCCGAGATTTGGCCAGAGCCGAGGTC 600  
DB 91971 AAGCAGAGCTTGTGAGACTTCAAGATCCGAGCCGAGATTTGGCCAGAGCCGAGGTC 92030  
QY 601 GGGCCCTATACACCGAGTTCTCATGATGATCAAGACAGAGAGAGAGAGAGAGAGAG 660  
DB 92031 GGGCCCTATACACCGAGTTCTCATGATGATCAAGACAGAGAGAGAGAGAGAGAGAG 92090  
QY 661 AAACCGGCTGTACTCTCAAGGCGCGCCGCAATCCGAGACACAGAGATGAGAGAC 720  
DB 92091 AAACCGGCTGTACTCTCAAGGCGCGCCGCAATCCGAGACACAGAGATGAGAGAC 92150  
QY 721 TTCATGAG 780  
DB 92151 TTCATGAG 92210  
QY 781 GAGTTCTGACGAGGAGCTTCTCGAGAGCTGACGAGCGGTACCAACGAGAGCGTCGAG 840  
DB 92211 GAGTTCTGACGAGGAGCTTCTCGAGAGCTGACGAGCGGTACCAACGAGAGCGTCGAG 92270  
QY 841 AACATGAG 900  
DB 92271 AACATGAG 92330  
QY 901 ATGAG 960  
DB 92331 ATGAG 92390  
QY 961 CGGTGCGGAGAGCTGAGAGCTGAGAGCTGAGAGCTGAGAGAGAGAGAGAGAGAGAGAG 1020  
DB 92391 CGGTGCGGAGAGCTGAGAGCTGAGAGCTGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 92450  
QY 1021 ACCGAG 1080  
DB 92451 ACCGAG 92510

RESULT 12  
BC022111 1488 bp mRNA linear ROD 16-APR-2003

LOCUS Mus musculus cardiac lineage protein 1, mRNA (CDNA clone MGC:36249  
DEFINITION IMAGE:5067620), complete cds.

BC022111  
VERSION BC022111.1 GI:18381039

KEYWORDS MGC.

SOURCE Mus musculus (house mouse)

ORGANISM

REFERENCE  
AUTHORS  
1 (bases 1 to 1488)  
Strausberg, R.L., Feingold, E.A., Gronow, L.H., Derge, J.G.,  
Klausner, R.D., Collins, F.S., Wagner, L., Shemmen, C.M., Schuler, G.D.,  
Altschul, S.F., Zeeberg, B., Buetow, K.H., Scheffer, C.F., Bhat, N.K.,  
Hopkins, R.F., Jordan, H., Moore, T., Wax, S.I., Wang, J., Hsieh, F.,  
Diatchenko, L., Marusina, K., Farmer, A.A., Rubin, G.M., Hong, L.,  
Stapleton, M., Soares, M.B., Bonaldo, M.F., Casavant, T.L.,  
Scheetz, T.E., Brownstein, M.J., Uesdin, T.B., Toshiyuki, S.,  
Carinci, P., Prange, C., Raha, S.S., Loquellano, N.A., Peters, G.J.,  
Abramson, R.D., Mulle, S.J., Bosak, S.A., McSwan, P.J.,  
McKernan, K.J., Malek, J.A., Gunaratne, P.H., Richards, S.,  
Worley, K.C., Hale, S., Garcia, A.M., Gay, L.J., Hui, Y.K., S.W.,



Villalon, D.K., Muzny, D.M., Sodergren, E.J., Lu, X., Gibbs, R.A., Fahney, J., Helton, E., Ketterman, M., Madan, A., Rodrigues, S., Sanchez, A., Whiting, M., Madan, A., Young, A.C., Shevchenko, Y., Bouffard, G.G., Blakesley, R.W., Touchman, J.W., Green, E.D., Dickson, M.C., Rodriguez, A.C., Grimwood, J., Schmutz, J., Myers, R.M., Butterfield, Y.S., Krzywinski, M.I., Skalska, U., Smalhus, D.E., Schermer, A., Schein, J.E., Jones, S.J. and Marra, M.A.  
Generation and initial analysis of more than 15,000 full-length human and mouse cDNA sequences  
Proc. Natl. Acad. Sci. U.S.A. 99 (26), 16899-16903 (2002)

JOURNAL MEDLINE 22388257  
PUBMED 12477932  
REFERENCE 2 (bases 1 to 1488)  
AUTHORS Strausberg, R.  
JOURNAL Direct Submission  
Submitted (24-JAN-2002) National Institutes of Health, Mammalian Gene Collection (MGC), Cancer Genomics Office, National Cancer Institute, 31 Center Drive, Room 11A03, Bethesda, MD 20892-2590, USA

REMARK NIH-MGC Project URL: <http://mgc.nci.nih.gov>  
COMMENT Contact: MGC help desk  
Email: [cgabs-remail.nih.gov](mailto:cgabs-remail.nih.gov)

Tissue Procurement: Gilbert Smith, Ph.D.  
cDNA Library Preparation: Life Technologies, Inc.  
cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LNL)  
DNA Sequencing by: Genome Sequence Centre,  
BC Cancer Agency, Vancouver, BC, Canada  
[info@bgsc.bc.ca](mailto:info@bgsc.bc.ca)  
Steven Jones, Jennifer Amano, Ian Bosdet, Yaron Butterfield, Susana Chan, Readman Chiu, Chris Fjell, Erin Garland, Ran Guin, Leticia Hsiao, Martin Krzywinski, Reta Kutsche, Oliver Lee, Soo Sen Lee, Victor Ling, Carrie Mathewson, Candice Mcleavy, Steven Ness, Pawan Pandoh, Anna-Lisa Prabhu, Parvaneh Saeedi, Jacqueline Schein, Duane Smalhus, Michael Smith, Lorraine Spence, Jeff Scott, Michael Thorne, Miranada Tsai, Natasja van den Bosch, Jill Vardy, George Yang, Scott Zuyderduyn, Marco Marra.

Clone distribution: MGC clone distribution information can be found through the I.M.A.G.E. Consortium/LNL at: <http://image.lnl.gov>  
Series: IRK Plate: 57 Row: a Column: 15  
This clone was selected for full length sequencing because it passed the following selection criteria: Hexamer frequency ORF analysis, similarity but not identity to protein.

FEATURES  
source Location/Qualifiers

1..1488  
/organism="Mus musculus"  
/mol\_type="mRNA"  
/strain="C57BL/6J"  
/db\_xref="taxon:10090"  
/clone="MGC:36249 IMAGE:5067620"  
/tissue\_type="mammary tumor metastasized to lung. Tumor arose spontaneously from a senescent normal mammary (clonal) outgrowth infected with the virus MMTV."  
/clone\_1lb="NCI CGAP Lu29"  
/lab\_host="DH10B"  
/note="Vector: pCMV-SPORT6"  
1..1488  
/gene="C1p1-pending"  
/note="synonyms: H1S1, CLP-1"  
/db\_xref="Locustid:192231"  
/db\_xref="MGI:2385923"  
448..1188  
/codon\_start=1  
/product="C1p1-pending protein"  
/protein\_id="AAH2111.1"  
/db\_xref="GI:18381040"  
/translation="MSDSLVPQHSSEATKQAPAPAGGPEWQOOLGKKRRRP  
SKRRHWKPYKLTWEKKKDFDKSLRVAAPAPAGOPVAPNTTLOPMDDHO  
EBSDLTKGLYPRKAAKSDTDEDEVEAGBEDGSDGMDGDSFFLRDSEYER  
YHAESLQNMKSOLIKELYLELEKCLSRKDENNRLKSLRGVDARVELLELDR  
LRAENLQLTLEMLHROQERAPLSKRGD"

BASE COUNT

385 a 371 c 449 g 283 t

# ORIGIN

Query Match 77.2%; Score 833.8; DB 10; Length 1488;  
Best Local Similarity 86.9%; Pred. No. 9.9e-129;  
Matches 932; Conservative 0; Mismatches 132; Indels 9; Gaps 1;

QY 8 AGCCATTCTTGTGACAAATATCAACACAGCCTCAAACTAGCAACTGTACAGGTGCTGCTG 67  
DB 125 ACCCACTTTGACAGAACATCAACACAGCCTCAAACTAGCAACTGTACAGGTGCTGCTG 184  
QY 68 CTGTCCAGGAAGAGCTGAACCTTGAGCCGCCCCAGGCGCGGAGAGAGCGGGTCCCGAG 127  
DB 185 TTGTCCATGAAAGAGATACCTCTGAGCGCCCAAGCGGAGAGAACCGGTGCCCAAG 244  
QY 128 AGACAGTAGTGGCAATGAGAGCGTTCCCAAGTGGGTGGCGTCCCGGAGCGGAGAG 187  
DB 245 AGACAGTAGTGGCAATGAGAGCGTTCCCAAGTGGGTGGCGTCCCGGAGCGGAGAG 304  
QY 188 GGGAAAGGAGCCTGGAATCCCAACCACTCCCTTCAGACCCAGGCGCTGTCCGAACTTA 247  
DB 305 GAGAAAGGAGCCTGGAATCCCAACCACTCCCTTCAGACCCAGGCGCTGTCCGAACTTA 364  
QY 248 GCTGCTGAGAGAGGCGGAGAGAGGCGCAGATGGGAGCGACTGCTCCGTGGCGGAGCT 307  
DB 365 GCTCCCTGAGAGAGGCGGAGAGAGGCGCAGATGGGAGCGACTGCTCCGTGGCGGAGCT 424  
QY 308 TCCGCGCGCGGAGAGAGGAGAGAGGAGAGGAGAGGAGAGGAGAGGAGAGGAGAGGAG 367  
DB 425 CCCCCTGCGGAG 475  
QY 368 ATGACTCCGAGGCGAGTAAAGTGGGGGCTCTGCGGAGAGAGAGAGAGAGAGAGAGAG 427  
DB 476 ATGACTCCGAGGCGAGTAAAGTGGGGGCTCTGCGGAGAGAGAGAGAGAGAGAGAGAG 535  
QY 428 AGCAG 487  
DB 536 AGCAG 595  
QY 488 ATTGAACCGGAGTAAAGTGGGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 547  
DB 596 ATTGAACCGGAGTAAAGTGGGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 655  
QY 548 GCTTCGAGCTTCAAGAGATCCGAGCGGAGATTTGCCAAGGAGCGAGCGGTGCGGCCCT 607  
DB 656 GCTTCGAGCTTCAAGAGATCCGAGCGGAGATTTGCCAAGGAGCGAGCGGTGCGGCCCT 715  
QY 608 ATTAACACACGAGATTTCTCATGATGATACGACCGAGAGAGCGGAGTCTCAAAACCG 667  
DB 716 ATTAACACACGAGATTTCTCATGATGATACGACCGAGAGAGCGGAGTCTCAAAACCG 775  
QY 668 GCGTGTACTCCAAAGGCGCGCGCGCAAAATCCGAGAGAGAGAGAGAGAGAGAGAGAG 727  
DB 776 GCGTGTACTCCAAAGGCGCGCGCGCAAAATCCGAGAGAGAGAGAGAGAGAGAGAGAG 835  
QY 728 AAGAAAGGAGTGAAGAGATGGGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 787  
DB 836 AAGAAAGGAGTGAAG 895  
QY 788 TCGACCGGAGCTTCTCGGAG 847  
DB 896 TCGACCGGAGCTTCTCGGAG 955  
QY 848 GGAACGAGAGAGTCAATCAAG 907  
DB 956 GGAACGAGAGAGTCAATCAAG 1015  
QY 908 AGCAAAACAACCGGCTGCGGCTGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 967  
DB 1016 AGCAAAACAACCGGCTGCGGCTGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1075  
QY 968 GGAACGAGAGTGAAG 1027  
DB 1076 GGAACGAGAGTGAAG 1135

QY 1028 ACGACTGCACCGGACGAGAGCGCGCTTTCCAGTTTGAGACTAG 1080  
 Db 1136 ACGACTGCACCGGACGAGAGCGCGCTTTCCAGTTTGAGACTAG 1188

RESULT 13  
 AC120950/c  
 LOCUS  
 DEFINITION  
 AC120950 213625 bp DNA 11near HTG 19-NOV-2002  
 Rattus norvegicus clone CH230-456H21, WORKING DRAFT SEQUENCE, 5  
 unordered pieces.  
 AC120950  
 AC120950.4 GI:25091749  
 HTG: HTGS\_PHASE1; HTGS\_DRAFT; HTGS\_FULLTOP.  
 Rattus norvegicus (Norway rat)  
 Rattus norvegicus  
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae;  
 Rattus.

REFERENCE  
 AUTHORS  
 1 (bases 1 to 213625)  
 Muzny, D., Marle, Metzger, M., Lee, A., Abramson, S., Adams, C., Alder, J.,  
 Allen, C., Allen, H., Albrooks, S., Amin, A., Anguiano, D.,  
 Anyalebechi, V., Aoyagi, A., Ayodeji, M., Baca, E., Baden, H.,  
 Baldwin, D., Bandaranaike, D., Barber, M., Barnstead, M., Benahmed, F.,  
 Biswal, K., Blair, J., Blankenburg, K., Blyth, P., Brown, M.,  
 Bryant, N., Bunay, C., Butcher, P., Butrell, K., Calderon, E.,  
 Cardenas, V., Carter, K., Cavazos, I., Caesar, H., Center, A.,  
 Chacko, J., Chavez, D., Chen, G., Chen, R., Chen, Y., Chen, Z., Chu, J.,  
 Cleveland, C., Cockrell, R., Cox, C., Coyle, M., Cree, A., D'Souza, L.,  
 Davila, M. L., Davis, C., Davy-Carroll, L., De Anda, C., Deederich, D.,  
 Delgado, O., Denson, S., Deramo, C., Ding, Y., Dinh, H., Divya, K.,  
 Draper, H., Dugan-Rocha, S., Dunn, A., Durbin, K., Duval, B., Evans, K.,  
 Egan, A., Escotto, M., Eugene, C., Evans, C. A., Falls, T., Fan, G.,  
 Fernandez, S., Finley, M., Flagg, N., Forbes, L., Foster, M., Foster, P.,  
 Frazer, C. M., Gabisi, A., Garcia, R., Garcia, A., Garner, T., Garza, M.,  
 Gebregeregs, E., Geer, K., Gill, R., Grady, M., Guerra, M., Guevara, W.,  
 Gunaratne, P., Haaland, W., Hamill, C., Hamilton, C., Hamilton, K.,  
 Hernandez, R., Hines, S., Hladun, S. L., Hodgson, A., Hogue, M.,  
 Harney, Y., Havlak, P., Hawes, A., Henderson, N., Hernandez, J.,  
 Hollins, B., Howells, S., Hulky, S., Hume, J., Idlebird, D., Jackson, A.,  
 Jackson, L., Jacob, L., Jiang, H., Johnson, B., Johnson, R., Jolivet, A.,  
 Karpach, S., Kelly, S., Kelly, S., Khan, Z., King, L., Kovar, C.,  
 Kows, C., Krat, C. L., Ledow, H., Levan, J., Lewis, L., Li, Z., Liu, J.,  
 Liu, J., Liu, W., Liu, Y., London, P., Longacre, S., Lopez, J.,  
 Lorensueta, L., Louised, H., Lozano, R. J., Lu, X., Ma, J.,  
 Maheshwari, M., Mahindartne, M., Mahmoud, M., Mallory, K., Mangum, A.,  
 Mangum, B., Mapa, P., Martin, K., Martin, R., Martinez, E.,  
 Mawhney, S., McLeod, M. P., McNeill, T. Z., Meenen, E.,  
 Milosavljevic, A., Miner, G., Ming, E., Montemayor, J., Moore, S.,  
 Morgan, M., Morris, K., Morris, S., Munitasa, M., Murphy, M., Nair, L.,  
 Nankervis, C., Neal, D., Newton, N., Nguyen, N., Norris, S.,  
 Nwackalameh, O., Okunolu, G., Olarunpugoon, A., Pal, S., Parks, K.,  
 Pasernak, S., Paul, H., Perez, A., Perez, L., Pfankuch, C.,  
 Plopper, F., Poldexter, A., Popovic, D., Primus, E., Pu, L., L.,  
 Pucio, M., Quito, J., Rachlin, E., Reeves, K., Regier, M. A., Reich, R.,  
 Reilly, B., Reilly, M., Ren, Y., Reuter, M., Richards, S., Riggs, F.,  
 Rivers, C., Rodley, T., Rojas, A., Rose, M., Rose, R., Ruiz, S. J.,  
 Sanders, W., Savary, G., Scherer, S., Scott, G., Shattman, S., Shen, H.,  
 Shetty, J., Shvartbeyn, A., Sisson, I., Sitter, C. D., Smits, D.,  
 Sneed, A., Sodergren, E., Song, X.-Z., Sorelle, R., Sosa, D.,  
 Steimle, M., Strong, R., Sutton, A., Svatek, A., Tabori, R., Taylor, C.,  
 Taylor, T., Thomas, N., Thomas, S., Tingey, A., Trejos, Z., Usman, K.,  
 Valdes, R., Vera, V., Villanueva, D., Walden, L., Walker, B., Wang, J.,  
 Wang, O., Wang, S., Warren, J., Warren, R., Wei, X., White, P.,  
 Williams, G., Willson, R., Wleczek, R., Wooden, H., Worley, K.,  
 Wright, D., Wright, R., Wu, J., Yakub, S., Yen, U., Yoon, H., Yoon, V.,  
 Yu, F., Zhang, U., Zhou, J., Zhou, J., Zhao, S., Dunn, D., von  
 Niederhausern, A., Weiss, R., Smith, D. R., Holt, R. A., Smith, H. O.,  
 Weinstein, G., and Gibbs, R. A.  
 Direct Submission  
 2 (bases 1 to 213625)  
 Worley, K. C.

TITLE  
 JOURNAL  
 REFERENCE  
 AUTHORS  
 JOURNAL  
 COMMENT

Direct Submission  
 Submitted (14-MAY-2002) Human Genome Sequencing Center, Department  
 of Molecular and Human Genetics, Baylor College of Medicine, One  
 Baylor Plaza, Houston, TX 77030, USA  
 3 (bases 1 to 213625)  
 Rat Genome Sequencing Consortium.  
 Direct Submission  
 Submitted (19-NOV-2002) Human Genome Sequencing Center, Department  
 of Molecular and Human Genetics, Baylor College of Medicine, One  
 Baylor Plaza, Houston, TX 77030, USA  
 On Nov 19, 2002 this sequence version replaced gi:23907976.  
 The sequence in this assembly is a combination of BAC based reads  
 and whole genome shotgun sequencing reads assembled using Atlas  
 (<http://www.hgsc.bcm.tmc.edu/projects/rat/>). Each contig described  
 in the feature table below represents a scaffold in the Atlas  
 assembly (a 'contig-scaffold'). Within each contig-scaffold,  
 individual sequence contigs are ordered and oriented, and separated  
 by sized gaps filled with Ns to the estimated size. The sequence  
 may extend beyond the ends of the clone and there may be sequence  
 contigs within a contig-scaffold that consist entirely of whole  
 genome shotgun sequence reads. Both end sequences and whole genome  
 shotgun sequence only contigs will be indicated in the feature  
 table.

----- Genome Center  
 Center: Baylor College of Medicine  
 Center code: BCM  
 Web site: <http://www.hgsc.bcm.tmc.edu/>  
 Contact: hgsc-help@bcm.tmc.edu  
 ----- Project Information  
 Center project name: GWRQ  
 Center clone name: CH230-456H21  
 ----- Summary Statistics  
 Assembly program: Phrap; version 0.990329  
 Consensus quality: 196084 bases at least Q40  
 Consensus quality: 197843 bases at least Q30  
 Consensus quality: 198926 bases at least Q20  
 Estimated insert size: 201713; sum-of-contigs estimation  
 Quality coverage: 7x in Q20 bases; sum-of-contigs estimation

\* NOTE: Estimated insert size may differ from sequence length  
 \* (see [http://www.hgsc.bcm.tmc.edu/docs/genbank\\_draft\\_data.html](http://www.hgsc.bcm.tmc.edu/docs/genbank_draft_data.html)).  
 \* NOTE: This is a 'working draft' sequence. It currently  
 \* consists of 5 contigs. The true order of the pieces  
 \* is not known and their order in this sequence record is  
 \* arbitrary. Gaps between the contigs are represented as  
 \* runs of N, but the exact sizes of the gaps are unknown.  
 \* This record will be updated with the finished sequence  
 \* as soon as it is available and the accession number will  
 \* be preserved.

1 58094: contig of 58094 bp in length  
 \* 58095 58194: gap of unknown length  
 \* 58195 139497: contig of 8103 bp in length  
 \* 139498 139597: gap of unknown length  
 \* 139598 208746: contig of 69149 bp in length  
 \* 208747 208846: gap of unknown length  
 \* 208847 210051: contig of 1205 bp in length  
 \* 210052 210152: gap of unknown length  
 \* 210152 213625: contig of 3474 bp in length.

FEATURES  
 source  
 1..213625  
 /organism="Rattus norvegicus"  
 /mol\_type="genomic DNA"  
 /db\_xref="taxon:10116"  
 /clone="CH230-456H21"  
 1..1190  
 /note="wgs end extension  
 clone\_end:5p6"  
 47586..47846  
 /note="clone boundary  
 clone\_end:3p6  
 site:""  
 end\_sequence:B2240544"  
 49325 c 49186 g 48940 t 14232 others

BASE COUNT 51942 a 49325 c 49186 g 48940 t 14232 others



## ORIGIN

Query Match 75.7%; Score 818; DB 2; Length 213625;  
 Best Local Similarity 86.2%; Pred. No. 2,1e-126;  
 Matches 931; Conservative 0; Mismatches 140; Indels 9; Gaps 2;

QY 1 ATGGCCGAGCCATTTCTTTCAGAAATATCAACACAGCTTCAAACTAGCACTGTACAGGT 60  
 DB 119733 ATGGCTAGAGCCATCTTGTGAGAAATCAACACAGCTTCAAACTAGCACTGTACAGGT 119674  
 QY 61 GCTGCTGCTGCTCCAGAGAGAGCTGAAACCTTGAGCGCCCCCAGCGCGAGAGAGCGGGT 120  
 DB 119673 GCTGCTGCTGCTCATAGAGAGAGAACTTGAGCGCCCCCAGCGCGAGAGAGCGGGT 119614  
 QY 121 CCCGAG 180  
 DB 119613 CTTAAG 119554  
 QY 181 CCGAG 240  
 DB 119553 CAG 119494  
 QY 241 GAATCTAGCTGCTGCTGAG 300  
 DB 119493 GAATCTAGCTGCTGCTGAG 119434  
 QY 301 GCGGAG 360  
 DB 119433 GGTGCT 119383  
 QY 361 CCTTGTCTAGTCT 420  
 DB 119382 CAGAGCTAGTCT 119323  
 QY 421 TGGGAG 480  
 DB 119322 TGGGAG 119263  
 QY 481 AAGCGGAG 540  
 DB 119262 AAGCGGAG 119203  
 QY 541 AAGCGGAG 600  
 DB 119202 AAGCGGAG 119143  
 QY 601 GCGGCTTATTAACACAG 660  
 DB 119142 GCGGCTTATTAACACAG 119083  
 QY 661 AAAACCGGCTTATTAACACAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 720  
 DB 119082 AAAACCGGCTTATTAACACAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 119023  
 QY 721 TTTATGAG 780  
 DB 119022 TTTATGAG 118963  
 QY 781 GAGTTTCTGAG 840  
 DB 118962 GAGTTTCTGAG 118903  
 QY 841 AACATGAG 900  
 DB 118902 AACATGAG 118843  
 QY 901 ATGAG 960  
 DB 118842 ATGAG 118783  
 QY 961 CGTGTGCGGAG 1020  
 DB 118782 CGTGTGCGGAG 118723

QY 1021 ACCGAG 1080  
 DB 118722 ACCGAG 118663

RESULT 14  
 AC136172/c  
 LOCUS  
 DEFINITION  
 Rattus norvegicus clone CH230-22602, \*\*\* SEQUENCING IN PROGRESS  
 \*\*\* 53 unordered pieces.  
 AC136172 111803 bp DNA linear HTG 30-OCT-2002  
 AC136172.1 GI:24421642  
 HTG: HTGS PHASE1  
 Rattus norvegicus (Norway rat)  
 Rattus norvegicus  
 Rattus norvegicus  
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae;  
 Rattus.

REFERENCE  
 1 (bases 1 to 111803)  
 Muzny D, Marie, Metzger M, Lee, Abramson S, Adams C, Alder J,  
 Allen C, Allen H, Alsbrooks S, Amin A, Anguiano D,  
 Anyaledechi V, Aoyagi A, Ayodeji M, Baca E, Baden H,  
 Baldwin D, Bandaranaike D, Barber M, Barnstead M, Benahmed F,  
 Biswal K, Blair J, Blankenburg K, Blyth P, Brown M,  
 Bryant N, Buhay C, Burch P, Burrell K, Calderon E,  
 Cardenas V, Carter K, Cavazos I, Caesar H, Center A,  
 Chacko J, Chavez D, Chen G, Chen R, Chen Y, Chen Z, Chu J,  
 Cleveland C, Cockrell R, Cox C, Coyle M, Cree A, D'Souza L,  
 Davila M, Davis C, Davy-Carroll L, De Anda C, Dederich D,  
 Delgado O, Denson S, Deramo C, Ding Y, Dinh H, Divya K,  
 Duper H, Dugan-Rocha S, Dunn A, Durbin K, Duval B, Eaves K,  
 Egan A, Escotto M, Eugene C, Evans C, Falls T, Fan G,  
 Fernandez S, Finley M, Flagg N, Forbes L, Foster M, Foster P,  
 Fraser C, Gabisi A, Ganta R, Garcia A, Garner T, Garza M,  
 Gebregeorgis E, Geer K, Gilly R, Grady M, Guerra M, Guevara W,  
 Gunaratne P, Haaland W, Hamill C, Hamilton C, Hamilton K,  
 Harvey Y, Havlak P, Hawes A, Henderson N, Hernandez J,  
 Hernandez R, Hines S, Hladun S, L, Hodgson A, Hogue M,  
 Hollins B, Howells S, Hulik S, Hune J, Idlebird D, Jackson A,  
 Jackson L, Jacob L, Jiang H, Johnson B, Johnson R, Jolivet A,  
 Karpathy S, Kelly S, Kelly S, Khan Z, King L, Kovar C,  
 Kowitz C, Kraft C, Lebow H, Levan J, Lewis L, Li Z, Liu J,  
 Liu J, Liu W, Liu Y, London P, Longacre S, Lopez J,  
 Lorenshewa L, Louisseg H, Lozdo R, Lu X, Ma J,  
 Maneshwari M, Mahindaratne M, Mamoud M, Martinez E,  
 Mangum B, Mapua P, Martin K, Martin R, Martinez E,  
 Mawhinney S, McLeod M, McNeill T, Meenen B, Milosavljevic A,  
 Miner G, Mintz E, Montemayor J, Moore S, Morgan M, Morris K,  
 Morris S, Mundasa M, Murphy M, Nair L, Nankervis C, Neal D,  
 Newton N, Nguyen N, Norris S, Nwankwelen O, Okunolu G,  
 Olariunpogon A, Pal S, Parks K, Paeternak S, Paul H,  
 Perez A, Perez L, Pfannkuch C, Plopper F, Poindexter A,  
 Popovic D, Primus E, Pu L, L, Puaso M, Quiroz J, Rachlin E,  
 Reeves K, Regier M, A, Reigh R, Reilly B, Reilly M, Ren Y,  
 Reuter M, Richards S, Riggs P, Rives C, Rodkey T, Rojas A,  
 Rose M, Rose R, Ruiz S, J, Sanders M, Savary G, Scherer S,  
 Scott G, Shatsman S, Shen H, Shetty U, Shvartsbeyn A,  
 Sisson I, Sitter C, D, Smajd D, Sneed A, Sodergren E,  
 Song X, Z, Sorrell R, Soza J, Steidle M, Strong R, Sutton A,  
 Svarek A, Tabor P, Taylor C, Taylor T, Thomas N, Thomas S,  
 Tingey A, Treise Z, Umani K, Valse R, Vera V, Villaseana D,  
 Walgren L, Walker B, Wang J, Wang Q, Wang S, Warren J,  
 Warren R, Wei X, White F, Williams G, Willson R, Wleczky R,  
 Wooden H, Worley K, Wright D, Wright R, Wu J, Yakub X,  
 Yen J, Yoon L, Yoon V, Yu F, Zhang D, Zhou J, Zhou X,  
 Zhao S, Dunn D, von Niederhausen A, Weiss R, Smith D, R,  
 Holt R, A, Smith H, O, Weinstein G, and Gibbs R, A.

TITLE  
 JOURNAL  
 REFERENCE  
 AUTHORS  
 TITLE  
 Unpublished  
 2 (bases 1 to 111803)  
 Rat Genome Sequencing Consortium.  
 Direct Submission



```

Db 14924 CCGAGATG-TCGCCAAGGCCAGCCGGTTCGCCCTAATAACACAGCACTTNTCATGG 14866
QY 632 ATGATCAGCAGCAGAGAGCCGATCTCAAAACCGGCTGTATCTCCAGCGGCGCG 691
Db 14865 ATATCATTATATCAGAGAGACCTGATCTCAAAACCGGCTTTCCTTAAGCGGCGAGCG 14806
QY 692 CCAATTCGACGACACCGAGATGAGACTTTCATGAGAGAGGGGGTGAAGAGATGGGG 751
Db 14805 CCAATTCGACGACACCGAGATGAGACTTTCATGAGAGAGGGGGTGAAGAGAGAG 14746
QY 752 GCGAGCATGGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 811
Db 14745 GCGAGCATGGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 14686
QY 812 ACAGAGCGGTACACACCGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 871
Db 14685 ACAGAGCGGTACACACCGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 14626
QY 872 ACCTGGAAGTGGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 931
Db 14625 ACCTGGAAGTGGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 14566
QY 932 AAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 991
Db 14565 AAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 14506
QY 992 GCGTGGCGCGCGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1051
Db 14505 GCGTGGCGCGCGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 14446
QY 1052 GAGCGCGCGCTTTCAGAGTGGAGAGACTAG 1080
Db 14445 GAGCGCGCTTTCAGAGTGGAGAGACTAG 14417

```

```

RESULT 15
BD058159 523 bp DNA linear PAT 27-AUG-2002
LOCUS Secreted expressed sequence tags (ESTs)
DEFINITION BD058159
ACCESSION BD058159.1 GI:22603765
VERSION JP 2001519666-A/14.
KEYWORDS
SOURCE
ORGANISM
Zea mays
Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae; PACAD
clade; Panicoideae; Andropogoneae; Zea.
1 (bases 1 to 523)
JACOBS,K., MCCOY,J.M., LAVALLIE,E.R., RACIE,L.A., MERBERG,D.,
TREACY,M., SPAULDING,V. and AGOSTINO,M.J.
Secreted expressed sequence tags (ESTs)
Patent: JP 2001519666-A 14 23-OCT-2001;
GENETICS INSTITUTE INC
PN JP 2001519666-A/14
PD 23-OCT-2001
PR 10-APR-1998 JP 1998543068
PR 10-APR-1997 US 08/835913
PI KENNETH JACOBS, JOHN M MCCOY, EDWARD R LAVALLIE, LISA A RACIE, PI
DAVID MERBERG,
PI MAURICE TREACY, VIKKI SPAULDING, MICHAEL J AGOSTINO PC
CI2N15/12, CI2N5/10, C07K14/47, C12Q1/68, A61K38/17 CC Strandedness:
Double;
CC Topology: linear;
FH Key Location/Qualifiers.

```

```

FEATURES
source Location/Qualifiers
1..523
/organism="Zea mays"
/mol_type="genomic DNA"
/db_xref="taxon:4577"

```

```

BASE COUNT 112 a 149 c 153 g 108 t 1 others
ORIGIN

```

```

Query Match 36.6%; Score 395.8; DB 6; Length 523;

```

```

Best Local Similarity 96.6%; Pred. No. 5,3e-56;
Matches 403; Conservative 1; Mismatches 13; Indels 0; Gaps 0;
QY 1 ATGGCCGAGCCATTCTTTCAGAAATATCAACACAGAGCTCAAACTAGCACTGACAGT 60
Db 104 ATGGCCGAGCCATTCTTTCAGAAATATCAACACAGAGCTCAAACTAGCACTGACAGT 163
QY 61 GCTGCTGCTGTCAGAGAGAGCTGAAACCTTGAGCGCCCTCCAGCGCGGAGAGAGCGG 120
Db 164 GCTGCTGCTGTCAGAGAGAGCTGAAACCTTGAGCGCCCTCCAGCGCGGAGAGAGCGG 223
QY 121 CCGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 180
Db 224 CCGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 283
QY 181 CCGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 240
Db 284 CCGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 343
QY 241 GAATCTAGCTGCTGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 300
Db 344 GAATCTAGCTGCTGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 403
QY 301 GCGGACTTCCCGCCGCGGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 360
Db 404 GCGGACTTCCCGCCGCGGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 463
QY 361 CCTGTGATGACTCCGAGGCAAGTAGTGGGGGCTCTGCGGACAGAGGAGAGAGAG 417
Db 464 CCTGTGATGACTCCGAGGCAAGTAGTGGGGGCTCTGCGGACAGAGGCTCTTGGCCCTCGAG 520

```

```

Search completed: February 5, 2004, 11:18:36
Job time : 4192 secs

```

THIS PAGE BLANK (uspto)

GenCore version 5.1.6  
Copyright (c) 1993 - 2004 Compugen Ltd.

OM protein - protein search, using sw model

Run on: February 5, 2004, 13:10:24 ; Search time 110 Seconds  
(without alignments)  
842.190 Million cell updates/sec

Title: US-09-972-758A-2  
Sequence: 1 MAEPLESEXOHQOTSNCCTG.....LTENELHROQERAPLSKFGD 359

Scoring table: BLOSUM62  
Gapop 10.0, Gapext 0.5

Searched: 830525 seqs, 258052604 residues  
Total number of hits satisfying chosen parameters: 830525

Minimum DB seq length: 0  
Maximum DB seq length: 200000000  
Post-processing: Minimum Match 0%  
Maximum Match 100%  
Listing first 45 summaries

Database: SPTREMBL\_23:\*

- 1: sp\_archaea:\*
- 2: sp\_bacteria:\*
- 3: sp\_fungi:\*
- 4: sp\_human:\*
- 5: sp\_invertebrate:\*
- 6: sp\_mammal:\*
- 7: sp\_mhc:\*
- 8: sp\_organelle:\*
- 9: sp\_phage:\*
- 10: sp\_plant:\*
- 11: sp\_rodent:\*
- 12: sp\_virus:\*
- 13: sp Vertebrate:\*
- 14: sp\_unclassified:\*
- 15: sp\_virus:\*
- 16: sp\_bacteriophage:\*
- 17: sp\_archaeal:\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match %	Length	ID	Description
1	1910	100.0	359	4_O94992	O94992 homo sapien
2	1587.5	83.1	356	11_O8R409	O8R409 mus musculus
3	1199	62.8	246	11_O8VDF7	O8VDF7 mus musculus
4	545.5	28.6	286	4_O96MH2	O96MH2 homo sapien
5	483.5	25.3	280	11_O9D4C7	O9D4C7 mus musculus
6	213	11.2	360	5_O8INP6	O8INP6 drosophila
7	212	11.1	360	5_O8ICM8	O8ICM8 drosophila
8	209	10.9	349	5_O9VFB2	O9VFB2 drosophila
9	150	7.9	1142	5_O8T6B4	O8T6B4 drosophila
10	148.5	7.8	992	4_O9NTH6	O9NTH6 homo sapien
11	147	7.7	1455	4_O9NTH6	O9NTH6 homo sapien
12	142.5	7.5	560	6_O9SLS7	O9SLS7 macaca fasc
13	141.5	7.4	1132	5_O9W475	O9W475 drosophila
14	140.5	7.4	443	5_O761S3	O761S3 periplaneta
15	140	7.3	609	17_O8TXA4	O8TXA4 methanopyru
16	140	7.3	714	10_O94DE3	O94DE3 oryza sativ

17	139	7.3	1781	4_O9UKX0	O9UKX0 homo sapien
18	139	7.3	1890	4_O9UKW3	O9UKW3 homo sapien
19	139	7.3	2073	4_O9UKW2	O9UKW2 homo sapien
20	138.5	7.3	695	11_O8C148	O8C148 mus musculus
21	138.5	7.3	944	11_O922S3	O922S3 mus musculus
22	138.5	7.3	1665	11_O8VIT1	O8VIT1 mus musculus
23	138	7.2	2072	4_O8WYB5	O8WYB5 homo sapien
24	137.5	7.2	483	6_O8WJG1	O8WJG1 bos taurus
25	137.5	7.2	540	11_O9EO53	O9EO53 mus musculus
26	136	7.1	399	10_O94IK2	O94IK2 solanum tub
27	135.5	7.1	530	4_O9UER6	O9UER6 homo sapien
28	135.5	7.1	1402	3_O8NIV6	O8NIV6 neorospira
29	135	7.1	540	4_O96B87	O96B87 homo sapien
30	135	7.1	554	4_O9H8Q2	O9H8Q2 homo sapien
31	135	7.1	555	4_O8N217	O8N217 homo sapien
32	135	7.1	602	3_O9HEL8	O9HEL8 neorospira
33	135	7.1	813	4_O9ULN3	O9ULN3 homo sapien
34	133	7.0	462	11_O9LIB7	O9LIB7 ratu norv
35	133	7.0	538	15_O9WJ4	O9WJ4 drosophila
36	133	7.0	1144	5_O9MJ7	O9MJ7 drosophila
37	133	7.0	1737	15_O92808	O92808 meloney mur
38	132.5	6.9	530	4_O75475	O75475 homo sapien
39	132.5	6.9	911	5_O8T852	O8T852 drosophila
40	132	6.9	528	4_O9H6U3	O9H6U3 homo sapien
41	132	6.9	528	4_O9H3P7	O9H3P7 homo sapien
42	132	6.9	528	4_O8IZC5	O8IZC5 homo sapien
43	132	6.9	2400	4_O8IWP2	O8IWP2 homo sapien
44	132	6.9	2416	4_O8IWP1	O8IWP1 homo sapien
45	132	6.9	2432	4_O8IWP0	O8IWP0 homo sapien

ALIGNMENTS

RESULT 1	O94992	PRELIMINARY;	PP: 359 AA.
ID	O94992		
AC	O94992		
DT	01-MAY-1999 (TREMBLrel. 10, created)		
DT	01-MAY-1999 (TREMBLrel. 10, last sequence update)		
DT	01-UN-2001 (TREMBLrel. 17, last annotation update)		
DE	HIS1 protein (HMBB-inducible).		
GN	HIS1.		
OS	Homo sapiens (Human).		
OC	Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;		
OC	Mammalia; Eutheria; Primates; Catarrhini; Homidae; Homo.		
OX	NCBI_TaxID=9606;		
RP	SEQUENCE FROM N.A.		
RA	Kushnarev, Nagasaki K., Kimura K., Ishikawa S., Maas N., Ogawa S.,		
RA	Yamaguchi K.,		
RT	"Cloning of HMBB-inducible transcript, HIS1, in human vascular smooth		
RT	muscle cells."		
RL	Submitted (DEC-1998) to the EMBL/GenBank/DBJ databases.		
RP	SEQUENCE FROM N.A.		
RC	TISSUE:lung;		
RA	Straussberg R.,		
RL	Submitted (APR-2001) to the EMBL/GenBank/DBJ databases.		
DR	EMBL; AB021179; BAA36166.1; -		
DR	EMBL; BC006460; AA006460.1; -		
SO	SEQUENCE 359 AA; 40623 MW; B12845C4E2595FF0 CRC64;		
Query Match	100.0%; Score 1910; DB 4; Length 359;		
Best Local Similarity	100.0%; Pred. No. 9.5e-132;		
Matches 359; Conservative	0; Mismatches 0; Indels 0; Gaps 0;		
OY	1 MAEPLESEXOHQOTSNCCTGAAAOEELNPERPPGAEEVPEEDSRMOSRAPOLGGRG 60		
DB	1 MAEPLESEXOHQOTSNCCTGAAAOEELNPERPPGAEEVPEEDSRMOSRAPOLGGRG 60		
OY	61 PEGESLSQPPPLQTOACPESSCLREGEKGGNGDSSAGDPPEPAVEPTPEALLAQ 120		

```

Db 61 PEGEGSLSEQPPPLQTOACPESSCLREGEKQNGDDSSAGGDFPPPAEVEPTPEALLAQ 120
Qy 121 PCHDSASXLGAPAAAGEEEMGQOQOLGKKKRRRPSKKRHWKPYKLTWEKKKPFDE 180
Db 121 PCHDSASXLGAPAAAGEEEMGQOQOLGKKKRRRPSKKRHWKPYKLTWEKKKPFDE 180
Qy 181 KOSLRASRIAEWFAKQOPVAPYNTTQFLMDHDHDOEPPDLKTGLYSKRAAKSDTSDDD 240
Db 181 KOSLRASRIAEWFAKQOPVAPYNTTQFLMDHDHDOEPPDLKTGLYSKRAAKSDTSDDD 240
Qy 241 FMEEGEEDGSGMGDGESEFLQRPSEFYERHYTHESLQMSKQELIKELYELEKCLSR 300
Db 241 FMEEGEEDGSGMGDGESEFLQRPSEFYERHYTHESLQMSKQELIKELYELEKCLSR 300
Qy 301 MEDENRRLRESRRLGDDARVARELELELDRLRAENLQLTTEMLHROQERAPLSKFGD 359
Db 301 MEDENRRLRESRRLGDDARVARELELELDRLRAENLQLTTEMLHROQERAPLSKFGD 359

```

## RESULT 2

```

Q8R409 PRELIMINARY; PRT; 356 AA.
AC Q8R409;
DT 01-JUN-2002 (T-EMBLrel. 21, Created)
DT 01-JUN-2002 (T-EMBLrel. 21, Last sequence update)
DT 01-MAR-2003 (T-EMBLrel. 23, Last annotation update)
DE Cardiac lineage protein 1.
GN CLP1.
OS Mus musculus (Mouse).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
OC NCBI_Taxid=10090;
RN NCB1
RP SEQUENCE FROM N.A.
RC STRAIN=129/SvJ.
RA Huang F., Wagner M., Siddiqui M.;
RT "Structure, expression, and functional characterization of the mouse
RT CLP-1 gene."
RL Submitted (MAR-2002) to the EMBL/GenBank/DBJ databases.
DR EMBL; AY090614; AM009026.1; -.
DR MGI; MGI:2385923; Clp1.
SQ SEQUENCE 356 AA; 40243 MW; 242DE7EE66BA293B CRC64;

```

Query Match 83.1%; Score 1587.5; DB 11; Length 356;  
 Best Local Similarity 85.8%; Pred. No. 3.5e-108;  
 Matches 308; Conservative 11; Mismatches 37; Indels 3; Gaps 2;

```

Qy 1 MAEPFLSEYOHQPTNSNCTGAALVQELNDEPRPGAEERVPEDSRWQSRAPQLGGRPG 60
Db 1 MAEPFLTEHQHQPTNSNCTGAALVQELNDEPRPGAEERVPEDSRWQSRAPQLGGRPG 60
Qy 61 PEGEGSLSEQPPPLQTOACPESSCLREGEKQNGDDSSAGGDFPPPAEVEPTPEALLAQ 120
Db 61 PEGEGSLSEQPPPLQTOACPESSCLREGEKQNGDDSSAGGDFPPPAEVEPTPEALLAQ 120
Qy 61 QEGEGSLKIQPLPLQTNACPESSLSLEKGEKQNGEDLSTGG-ASFAEGEPMSSES--LVQ 117
Db 61 QEGEGSLKIQPLPLQTNACPESSLSLEKGEKQNGEDLSTGG-ASFAEGEPMSSES--LVQ 117
Qy 121 PCHDSASXLGAPAAAGEEEMGQOQOLGKKKRRRPSKKRHWKPYKLTWEKKKPFDE 180
Db 121 PCHDSASXLGAPAAAGEEEMGQOQOLGKKKRRRPSKKRHWKPYKLTWEKKKPFDE 180
Qy 118 PGHDSERATKOEARPAAGEEPWGOQROLGKKKRRRPSKKRHWKPYKLTWEKKKPFDE 177
Db 118 PGHDSERATKOEARPAAGEEPWGOQROLGKKKRRRPSKKRHWKPYKLTWEKKKPFDE 177
Qy 181 KOSLRASRIAEWFAKQOPVAPYNTTQFLMDHDHDOEPPDLKTGLYSKRAAKSDTSDDD 240
Db 181 KOSLRASRIAEWFAKQOPVAPYNTTQFLMDHDHDOEPPDLKTGLYSKRAAKSDTSDDD 240
Qy 178 KOSLRASRIAEWFAKQOPVAPYNTTQFLMDHDHDOEPPDLKTGLYPRRAAKSDTSDDD 237
Db 178 KOSLRASRIAEWFAKQOPVAPYNTTQFLMDHDHDOEPPDLKTGLYPRRAAKSDTSDDD 237
Qy 241 FMEEGEEDGSGMGDGESEFLQRPSEFYERHYTHESLQMSKQELIKELYELEKCLSR 300
Db 241 FMEEGEEDGSGMGDGESEFLQRPSEFYERHYTHESLQMSKQELIKELYELEKCLSR 300
Qy 238 FVEEAGEEDGSGMGDGESEFLQRPSEFYERHYTHESLQMSKQELIKELYELEKCLSR 297
Db 238 FVEEAGEEDGSGMGDGESEFLQRPSEFYERHYTHESLQMSKQELIKELYELEKCLSR 297
Qy 301 MEDENRRLRESRRLGDDARVARELELELDRLRAENLQLTTEMLHROQERAPLSKFGD 359
Db 301 MEDENRRLRESRRLGDDARVARELELELDRLRAENLQLTTEMLHROQERAPLSKFGD 359
Qy 298 KEDENRRLRESRRLGDDARVARELELELDRLRAENLQLTTEMLHROQERAPLSKFGD 356
Db 298 KEDENRRLRESRRLGDDARVARELELELDRLRAENLQLTTEMLHROQERAPLSKFGD 356

```

## RESULT 3

```

Q8VDF7 PRELIMINARY; PRT; 246 AA.
ID Q8VDF7;
AC Q8VDF7;
DT 01-MAR-2002 (T-EMBLrel. 20, Created)
DT 01-MAR-2002 (T-EMBLrel. 20, Last sequence update)
DT 01-MAR-2003 (T-EMBLrel. 23, Last annotation update)
DE Similar to HMBa-inducible.
GN CLP1.
OS Mus musculus (Mouse).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
OC NCBI_Taxid=10090;
RN NCB1
RP SEQUENCE FROM N.A.
RC Strausberg R.;
RA Submitted (JAN-2002) to the EMBL/GenBank/DBJ databases.
DR EMBL; BC022111; AAH2211.1; -.
DR MGI; MGI:2385923; Clp1.
SQ SEQUENCE 246 AA; 28572 MW; 6CF8227F599B77EF CRC64;

```

Query Match 62.8%; Score 1199; DB 11; Length 246;  
 Best Local Similarity 93.5%; Pred. No. 5.5e-80;  
 Matches 229; Conservative 5; Mismatches 11; Indels 0; Gaps 0;

```

Qy 115 ABLAQPCCHDSASXLGAPAAAGEEEMGQOQOLGKKKRRRPSKKRHWKPYKLTWE 174
Db 2 SBLVQPGHDSERATKOEARPAAGEEPWGOQROLGKKKRRRPSKKRHWKPYKLTWE 61
Qy 175 KKKPEKOSLRASRIAEWFAKQOPVAPYNTTQFLMDHDHDOEPPDLKTGLYSKRAAKSD 234
Db 62 KKKPEKOSLRASRIAEWFAKQOPVAPYNTTQFLMDHDHDOEPPDLKTGLYPRRAAKSD 121
Qy 235 DTSDDFMEEGEEDGSGMGDGESEFLQRPSEFYERHYTHESLQMSKQELIKELYEL 294
Db 122 DTSDDFMEEGEEDGSGMGDGESEFLQRPSEFYERHYTHESLQMSKQELIKELYEL 181
Qy 295 EKCLSRMEDENRRLRESRRLGDDARVARELELELDRLRAENLQLTTEMLHROQERAP 354
Db 182 EKCLSRMEDENRRLRESRRLGDDARVARELELELDRLRAENLQLTTEMLHROQERAP 241
Qy 355 SKFGD 359
Db 242 SKFGD 246

```

## RESULT 4

```

Q96MH2 PRELIMINARY; PRT; 286 AA.
ID Q96MH2;
AC Q96MH2;
DT 01-DEC-2001 (T-EMBLrel. 19, Created)
DT 01-DEC-2001 (T-EMBLrel. 19, Last sequence update)
DT 01-OCT-2002 (T-EMBLrel. 22, Last annotation update)
DE Hypothetical protein FLJ32384 (Similar to putative).
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OC NCBI_Taxid=9606;
RN NCB1
RP SEQUENCE FROM N.A.
RC Tissue=Skeletal muscle;
RA Oshima A., Takahashi-Fujii A., Tanase T., Imose N., Takeuchi K.,
RA Arita M., Mitsuhashi K., Yuuki H., Hara H., Sugiyama T., Irie R.,
RA Otsuki T., Sato H., Wakamatsu A., Ishii S., Yamamoto J., Isono Y.,
RA Kawai-Hiro Y., Saito K., Nishikawa T., Kimura K., Yamashita H.,
RA Matsuo K., Nakamura Y., Sekine M., Kikuchi H., Kanda K., Wagatsuma M.,
RA Muraoka K., Kanehori K., Sugiyama A., Kawakami B., Suzuki Y.,
RA Sugano S., Nagahara K., Masuho Y., Nagai K., Isegai T.;
RT "NEO human cDNA sequencing project."
RL Submitted (OCT-2001) to the EMBL/GenBank/DBJ databases.
RN NCB1
RP SEQUENCE FROM N.A.
RC Tissue=Eye;
RA Strausberg R.;

```

Submitted (MAR-2002) to the EMBL/GenBank/DBJ databases.  
 DR EMBL: AC056946; BAB71319.1; -  
 DR EMBL: BC025970; AAB25970.1; -  
 KW Hypothetical protein.  
 SQ SEQUENCE 286 AA; 32418 MW; 58576D72096A8A6F CRC64;

Query Match 28.6%; Score 545.5; DB 4; Length 286;  
 Best Local Similarity 43.8%; Pred. No. 3.7e-32;  
 Matches 128; Conservative 36; Mismatches 93; Indels 35; Gaps 6;

73 PIOTACPSSTCRBEKQNGDSDS-----SAGDFPPPAVPTPEAELLQPC-- 122  
 DB 5 PNOTACNAPSVALPEAKTSGAPSPQTPPERHDSGSLPFRPMHSHEDELLAGAVG 64  
 QY 123 --HDEASRLGAPPAAGEEEMGQOQOLGKKRRRPPSKKKHMKRYLLTWBEKKPDE 180  
 DB 65 LGMNSGSPPTQSBGSGSA-----AVLARKKRRRPPSKKKHMKRYLLTSLNAEQGRB 118  
 QY 181 KOSLRASRIRAEWPAKQGPVAPYNTTQFLMDHDOEP--DKTGLYSKRAAKSDDTSD 238  
 DB 119 ROSQRASRVREEMFAKQGPVAPYNTTQFLMDHDOEP--DKTGLYSKRAAKSDDTSD 173  
 QY 239 DPFMEGGEEDGSDGMDGDFLORDPSEYERHTHTSLQNMKQELIKYLEKCL 298  
 DB 174 -----EAGSDGRGAHGFQKDPSEYERHTHTSLQNMKQELIKYLEKCL 224  
 QY 299 SMWENNRLRLSKRLGDDAR-VRELELDRLAEHLQLTENELRQ 349  
 DB 225 SQAEETRLQOLQACTGQOSCHQVEBLAEVRLTEHRLQENQNMRE 276

RESULT 5

09DAC7 PRELIMINARY; PRT; 280 AA.

AC 09DAC7; 01-JUN-2001 (TREMBlrel. 17, Created)  
 DT 01-JUN-2001 (TREMBlrel. 17, Last sequence update)  
 DT 01-JUN-2001 (TREMBlrel. 17, Last annotation update)  
 DE 4933402121Rik protein.  
 GN 4933402121Rik.  
 OS Mus musculus (Mouse).  
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.  
 NCBI\_Taxid=10090;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=C57BL/6J; TISSUE=Testis;  
 RX MEDLINE=21085660; PubMed=11217851;  
 RA Kawai J., Shingawa A., Shibata K., Yoshino M., Itoh M., Ishii Y.,  
 RA Arakawa T., Hara A., Fukunishi Y., Kono H., Adachi J., Fukuda S.,  
 RA Aizawa K., Izawa M., Nishi K., Kiyosawa H., Kondo S., Yamana I.,  
 RA Saito T., Okazaki Y., Gojobori T., Bono H., Kasubawa T., Saito R.,  
 RA Kadoya K., Matsuda H.A., Ashburner M., Batalov S., Casavant T.,  
 RA Fleischman W., Gaasterland T., Gissi C., King B., Kochia H.,  
 RA Kuehl P., Lewis S., Matsuo Y., Nikaido I., Pesole G., Quackenbush J.,  
 RA Kuehl L.M., Staudl F., Suzuki R., Tomita M., Wagner L., Washio T.,  
 RA Sakai K., Okido T., Furuno M., Aono H., Balderelli R., Barsh G.,  
 RA Blake J., Boffelli D., Bojunga N., Catranci P., de Bonaldo M.F.,  
 RA Brownstein M.J., Bult C., Fletcher C., Fujita M., Gabori D.M.,  
 RA Gustinich S., Hill D., Hofmann M., Hume D.A., Kamuya M., Lee N.H.,  
 RA Lyons P., Marchionni L., Mashima J., Mazzarelli J., Mombaerts P.,  
 RA Nordone P., Ring B., Rindwald M., Rodriguez I., Sakamoto N.,  
 RA Sasaki H., Sato K., Schoenbach C., Seya T., Shibata Y., Storch K.-F.,  
 RA Suzuki H., Toyokawa K., Wang K.H., Weltz C., Whitaker C., Wilming L.,  
 RA Wynshaw-Boris A., Yoshida K., Hasegawa Y., Kawaji H., Kohsaki S.,  
 RA Hayashizaki Y.;  
 RT "Functional annotation of a full-length mouse cDNA collection.";  
 RL Nature 409:685-690(2001).  
 DR EMBL: AK016624; BAB30344.1; -  
 DR MED: MGI:1918309; 4933402121Rik.  
 SQ SEQUENCE 280 AA; 32388 MW; 6BA925FE41A6334 CRC64;

Query Match 25.3%; Score 483.5; DB 11; Length 280;

Best Local Similarity 44.4%; Pred. No. 1.2e-27;  
 Matches 112; Conservative 29; Mismatches 66; Indels 45; Gaps 7;

122 CHDSBASK-----LGA--PAAGEEEMGQOQOLGKKRRRPPSKKKRH 163  
 DB 12 CHQDRRSORNRKGLSPLMAQVYTGVLPEPAQAVNR-----RCMPERSTYGALEAQRH 67  
 QY 164 WKPYKLTWEKKKPKDEKOSLRASRIRAEWPAKQGPVAPYNTTQFLMDHDOEPDLTKG 223  
 DB 68 WRPYELTSLNAEQGRB-----AVLARKKRRRPPSKKKHMKRYLLTSLNAEQGRB 124  
 QY 224 LYSKRAAKSDDTSDDPMEEGE-EDGSDGMDGDFLORDPSEYERHTHTSLQNM 282  
 DB 125 -----DVLHPSHSGSGENEAQSDGQGAHGFQKDPSEYERHTHTSLQNR 174  
 QY 283 SKQELIKYLEKCLSKRMEENNRLRLSKRLGDDAR-VRELELDRLAEHLQLTENELRQ 337  
 DB 175 SKQELVRYLDLRLSQAEQETRLR-----OLQCSRPQOQVEBLAEVRLTEHRLQENQ 230  
 QY 338 QLTENELRQ 349  
 DB 231 RLQENQNMRE 242

RESULT 6

08INF6 PRELIMINARY; PRT; 360 AA.

AC 08INF6; 01-MAR-2003 (TREMBlrel. 23, Created)  
 DT 01-MAR-2003 (TREMBlrel. 23, Last sequence update)  
 DT 01-MAR-2003 (TREMBlrel. 23, Last annotation update)  
 DE CG3508-PB.  
 GN CG3508.  
 OS Drosophila melanogaster (Fruit fly).  
 OC Eukaryota; Metazoa; Arthropoda; Hexapoda; Insecta; Pterygota;  
 OC Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;  
 OC Ephydroidea; Drosophilidae; Drosophila.  
 NCBI\_Taxid=7227;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC MEDLINE=20196006; PubMed=10731132;  
 RA Adams M.D., Celniker S.B., Holt R.A., Evans C.A., Gocayne J.D.,  
 RA Amanatides P.G., Scherer S.E., Li P.W., Hoiseth R.A., Galie R.F.,  
 RA George R.A., Lewis S.E., Richards S., Ashburner M., Henderson S.N.,  
 RA Sutton G.G., Mortman J.R., Yandell M.D., Zhang Q., Chen L.X.,  
 RA Brandon R.C., Rogers Y.H., Blazer R.G., Champ M., Pfeiffer B.D.,  
 RA Wan K.H., Doyle C., Baxter B.G., Helt G., Nelson C.R., Gabor G.L.,  
 RA Abail J.F., Agbayani A., An H.J., Andrews-Pfannkoch C., Baldwin D.,  
 RA Ballew R.M., Basu A., Baxendale J., Bayraktargolu L., Beasley E.M.,  
 RA Beeson K.V., Benos P.V., Berman B.P., Bhargava D., Bolshakov S.,  
 RA Borhova D., Botchan M.R., Bouck J., Brockstein P., Brothier P.,  
 RA Burris K.C., Busam D.A., Butler H., Cadieu E., Center A., Chandra I.,  
 RA Cherry J.M., Cawley S., Dahlke C., Davenport L.B., Davies P.,  
 RA de Pablos B., Delcher A., Deng Z., Mays A.D., Dew I., Dietz S.M.,  
 RA Dodson J.R., Douc L.E., Downes M., Dugan-Rocha S., Dunkov B.C., Dunn P.,  
 RA Durkin K.J., Evangelista C.C., Ferraz C., Ferreira S., Fleischmann W.,  
 RA Foaier C., Gabrielian A.E., Gang N.S., Gelbart W.M., Glaeser K.,  
 RA Glodex A., Gong F., Gorrell J.H., Gu Z., Guan P., Harris M.,  
 RA Harris N.L., Harvey D., Heiman T.J., Hernandez J.R., Houck J.,  
 RA Hostin D., Houston K.A., Howland T.J., Wei M.H., Ibegam C.,  
 RA Jalali M., Kaulush F., Karpen G.H., Ke Z., Kennan J.A., Ketchum K.A.,  
 RA Kimbel B.E., Kodira C.D., Kraft C., Kravitz S., Kulp D., Lai Z.,  
 RA Lasko P., Lei Y., Levitsky A.A., Li J., Li Z., Liang Y., Lin X.,  
 RA Liu X., Mattei B., McIntosh T.C., McLeod M.P., McPherson D.,  
 RA Merulov G., Milshina N.V., Mobarry C., Morris J., Moshrefi A.,  
 RA Mount S.M., Moy M., Murphy B., Murphy L., Muzny D.M., Nelson D.L.,  
 RA Nelson D.R., Nelson K.A., Nixon K., Nuskern D.R., Paclet J.M.,  
 RA Palazzolo M., Pittman G.S., Pan S., Pollard J., Puri V., Reese M.G.,  
 RA Reinert K., Remington K., Sanders R.D., Scheeler P., Shen H.,  
 RA Shue B.C., Siden-Kiamos I., Simpson M., Skupski M.P., Smith T.,  
 RA Spier E., Spirdling A.C., Stapleton M., Strong R., Sun E.,  
 RA Svirskas R., Tector C., Turner R., Venter E., Wang A.H., Wang X.,  
 RA Wang Z.Y., Wasmann D.A., Weinstein G.M., Weisenbach J.,

```

RA Williams S.M., Woodger, Morley K.C., Wu D., Yang S., Yao Q.A., Ye J.,
RA Yeh R.F., Zaveri J.S., Zhan M., Zhang G., Zhao Q., Zheng L.,
RA Zheng X.H., Zhong F.N., Zhong W., Zhou X., Zhu S., Zhu X., Smith H.O.,
RA Gibbs R.A., Myers E.W., Rubin G.M., Venter J.C.;
RT "The genome sequence of Drosophila melanogaster.";
RL Science 287:2185-2195 (2000).
RN [2]
RP SEQUENCE FROM N.A.
RA Celniker S.E., Adams M.D., Kronmiller B., Wan K.H., Holt R.A.,
RA Evans C.A., Gocayne J.D., Amanatides P.G., Brannon R.C., Rogers Y.,
RA Barton J., An H., Baldwin D., Benson J., Beeson K.Y., Buesam D.A.,
RA Carlson J.W., Center A., Chape M., Davenport L.B., Dietz S.M.,
RA Dodson K., Dorsett V., Doup L.E., Doyle C., Dresner K., Farfan D.,
RA Ferreira S., Frise E., Galle R.F., Garg N.S., George R.A.,
RA Gonzalez M., Houch J., Hoskins R.A., Hostin D., Howland T.J.,
RA Ibegwam C., Jalali M., Kruse D., Li P., Mattei B., Moshrefi A.,
RA McIntosh T.C., Moy M., Murphy B., Nelson C., Nelson K.A., Munro J.,
RA Patel J., Pargass V., Park S., Patel S., Pfeiffer B.,
RA Phanenavong S., Plittman G.S., Puri V., Richards S., Scheeler F.,
RA Stapleton M., Strong R., Sztrakeas R., Tector C., Tyler D.,
RA Williams S.M., Zaveri J.S., Smith H.O., Venter J.C., Rubin G.M.;
RT "Sequencing of Drosophila melanogaster genome.";
RL Submitted (MAR-2000) to the EMBL/GenBank/DBJ databases.
RN [3]
RP SEQUENCE FROM N.A.
RA Mira S., Crosby M.A., Matthews B.B., Bayraktaroglu I., Campbell K.,
RA Hradecky P., Huang Y., Kaminker J.S., Prochuk S.E., Smith C.D.,
RA Tupy J.L., Bergman C., Bernan B., Carlson J.W., Celniker S.E.,
RA Clamp M., Drysdale R., Emmert D., Frise E., de Grey A., Harris N.,
RA Kronmiller B., Marshall B., Millburn G., Richter J., Russo S.,
RA Searle S.M.J., Smith E., Shu S., Smutnick F., Whitfield E.,
RA Ashburner M., Gelbart W.M., Rubin G.M., Mungall C.J., Lewis S.E.;
RA "Annotation of Drosophila melanogaster genome.";
RL Submitted (MAR-2000) to the EMBL/GenBank/DBJ databases.
RN [4]
RP SEQUENCE FROM N.A.
RA Adams M.D., Celniker S.E., Gibbs R.A., Rubin G.M., Venter C.J.;
RL Submitted (MAR-2000) to the EMBL/GenBank/DBJ databases.
RN [5]
RP SEQUENCE FROM N.A.
RA FlyBase;
RA Submitted (SEP-2002) to the EMBL/GenBank/DBJ databases.
DR EMBL, AE003705, AA013614.1,
SQ SEQUENCE 360 AA; 39211 MW; 41F350CA35965233 CRC64;
Query Match 11.2%; Score 213; DB 5; Length 360;
Best Local Similarity 27.8%; Pred. No. 9.6e-08;
Matches 69; Conservative 38; Mismatches 73; Indels 68; Gaps 10;
QY 125 SEASHTGAPAAAG--EEEWGQOORQLG-----KKKRR----- 155
DB 2 AEAVNGNGNRHSGSAEKESGSGQRPLDGGGGGAGGGGVAAGGSGMPKRGKRGKSK 61
QY 156 -RPSKKRHHMPYYLTWE----EKKRDEKQSLASRIIRAMFPAKQPVAPYNTQRLM 210
DB 62 WQPKTKKXHY-PQWKLDMSTGAGATLEGNQNSTKLVRSSTL-----LVYNTNRFLLM 115
QY 211 DDHOQEBPDLKTGLYSKRAAKSDPTSDDDFMEEBEGEEDGSDGSGGSEFLQDFSET 270
DB 116 EHNSE-----LHKDSDDNCF--GSQTB-----DQVFLSKFSQDV 150
QY 271 YERYHTSLQNNMSKQELIKELYLEKLSRMEDENNRLRLSKRLGDDARVLELELD 330
DB 151 YERARLERLETMSKQELIQECMQIQRDRYSKQONISKEP--CAKLRAQDDKIRQLSRENO 207
QY 331 RLRAENLQ 338
DB 208 FLRTHLRL 215

```

[illegible]



```

Query Match      10.9%; Score 209; DB 5; Length 349;
Best Local Similarity 25.9%; Pred No. 1,8e-07;
Matches 68; Conservative 43; Mismatches 84; Indels 68; Gaps 10.

Oy      85 LREGEKGONCD-----DSAGGDFPPPAVEPTPEABLLAOPCHDSEASKUGAPAAAGEE 139
Db      1 MAEAIVKNSSGGQRPRLDSGGGG--                               -ASGGGVAVVGGS 36

Oy     140 EWGQQOROLGKKKRRRPSKKKKRHMKPYKLTFW---EKKKFDEKOSLRASRIAEHPA 195
Db      37 GMPKPKHRRGKKS-KMQPCKTKNHY-PWMKLDMSGAGATLEGONROMSRTQLVRSRL- 93

Oy     196 KGQPAVPANTQFLMDDHQEBPDILKTGLYSKRAAKSPDTSDDFMEEGEGEDGGSGQM 255
Db      94 ----LVPTNNRFLEEHNSF-----LHKODSDNCF----GSOTE----- 126

Oy     256 GGDGSEFIORDFSERYRYHTESLONMSKOELIKELYLELEKLSHMEEDENNRLRESRL 315
Db     127 --DYLFISKERSDYERARLERLETWSGQBELICGMQIIEBRYSQAQNISKEF---GAKL 181

Oy     316 GGDDARVRBLEJELDRLEAENIQ 338
Db    182 RAODDKIROLSRENQFLRTLHLR 204

RESULT 9
08T6B4
```

ID	Q9T6B4	PRELIMINARY;	PRT;	1142 AA.
DT	01-JUN-2002 (Tremblrel. 21, Created)			
DT	01-JUN-2002 (Tremblrel. 21, Last sequence update)			
DT	01-MAR-2003 (Tremblrel. 23, Last annotation update)			
DE	Non-transporter ABC protein AbcF4.			
GN	ABCf4.			
OS	Eukaryota; Mycetozoa; Dictyostellida; Dictyostellium.			
OC	Eukaryota; Mycetozoa; Dictyostellida; Dictyostellium.			
OX	NCBI_TaxID=44689;			
RN	SEQUENCE FROM N.A.			
RP	STRAIN=Ax4;			
RC	Aojard C., Loomis W.F.;			
RT	"Evolution of the ABC transporters of Dictyostellium."			
RL	Submitted (Feb-2002) to the EMBL/Genbank/DBJ databases.			
DR	EMBL; AF479256; AAL87694.1; -			
DR	InterPro; IPR003593; AAA_ATPase.			
DR	InterPro; IPR003439; ABC_transporter.			
DR	Pfam; PF00005; ABC_tran; 2.			
DR	ProDom; PD00006; ABC_transporter; 2.			
DR	SMART; SMO0382; AAA; 2.			
DR	PROSITE; PS00211; ABC_TRANSPORTER; 2.			
KW	ATP-binding.			
SQ	SEQUENCE 1142 AA; 130224 MW; 33B1815AB0942DC CRC64;			
Query Match	7.9%; Score 150; DB 5; Length 1142;			
Best Local Similarity	19.8%; Pred. No. 0.015;			
Matches	79; Conservative 173; Mismatches 132; Indels 114; Gaps 18;			
Qy	30 PERPAGAEHYRPEDSRWQSAAPQ---LGRPGPEEGSLSEQ-----PPPLQTGA 78			
Db	132 PQKGGKQQQ--QQDSDEQEIIPQVKKGGKAPQKGGKQDSDDEDEIPQVKKGG 189			
Qy	79 CPSSSCLREGKQNGDSSAGQFPPEA-----EVEPTPEALLAQPC 122			
Db	190 KPAQ--KKGGKQDSEDEDEDEVDQPVKKGNKKKKGVKHYEHEEHEEHEEHEEIPV 247			
Qy	123 HDSEASFLGAPAG-----EEEWGQQQLGKKGRRRPPSKKKRWKPYKLTWE 173			
Db	248 --KKGAKAPPKPKKGGKSKQSEDEEDVDQPVKKGGKKDKKKSKHY-----E 294			
Qy	174 EKKKFDKQSLR-----ASIRAEEMAKQAPVAPYTTQQLMDHQQEEDLTGL--- 224			
Db	295 EEEHEEHEEHEEIPQVKKGSNNKQKKGGKGRVVE-----EEHEEHEEHEEIPVKVG 346			
Qy	225 -----YSKRAAKSDDTSDDDPWE-----EGGEDGSDGMDGDSFLLQRFSETYR 273			
Db	347 SNKKDQKKGGKQKQDSEDEDEEIPQPVKKGGKKDKK-----GSHVVEHEEHEEHEE 400			
Qy	274 YHTESLQMSKQELIKYELKLSRMEDENNRLRESKRLGSDDARVRELELE--- 328			
Db	401 EEIE-----QVKKGGKKDK--KSSLDQMSLSISKK--GKGKGVHEEHEEHEE 450			
Qy	329 -----LDRLAENLQLITNELHROQRAPLSK 356			
Db	451 EEKPKSKNNKDKKKGKRV--EEHEEHEEHEEHEEKP 486			
RESULT 10				
Q9NTH6	PRELIMINARY;	PRT;	992 AA.	
Q9NTH6				
DT	01-OCT-2000 (Tremblrel. 15, Created)			
DT	01-OCT-2000 (Tremblrel. 15, Last sequence update)			
DT	01-OCT-2002 (Tremblrel. 22, Last annotation update)			
DE	Hypothetical protein (Fragment).			
GN	DKFZP43O2413.			
OS	Homo sapiens (Human).			
OC	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;			
OC	Mammalia; Eutheria; Primates; Catarrhini; Hominiidae; Homo.			
OX	NCBI_TaxID=9606;			
RN	[1]			

RP SEQUENCE FROM N.A.  
 RC TISSUE=Testis;  
 RA Koehler K., Beyer A., Mewes H.W., Gassenhuber J., Wiemann S.;  
 RL Submitted (JAN-2000) to the EMBL/GenBank/DBJ databases.  
 DR EMBL: AL137265; CAB70664.1; -  
 KM Hypothetical protein.  
 FT NON\_TER 1  
 SQ SEQUENCE 992 AA; 112628 MW; 039DF5B1E97E02F CRC64;  
 Query Match 7.8%; Score 148.5; DB 4; Length 992;  
 Best Local Similarity 21.1%; Pred. No. 0.016;  
 Matches 94; Conservative 66; Mismatches 149; Indels 137; Gaps 17;  
 QY 25 QEELNRPFGAEEVP-----EEDSRWQSRF-----POLGPRGEGE-----64  
 DB 230 QDELQSKQSGLEERHRLSPPLPHEERAOQSPRLATEEPPQ-----GREGQPEWKEAEL 284  
 QY 65 -----GSLSEQPPPLQTA-CPESSCLREGEKQNGDSSAGDFFPPAE-----VEPTP- 113  
 DB 285 GEDSAASLSTLQSLQREQAPSPPAACEKQKQHSQAELGPGQEAEDPEEKVAVSPTPP 344  
 QY 114 -----EALLAQPCHDSEASKLGPAPAGSEEWQGOQROLGKK-----152  
 DB 345 VSPERVSTEVAPPEQLSEAA-LKAMEAVAVQLEQDRHLLSKQKQKQQLREKLQEE 403  
 QY 153 -----HRRRPSKKRHKMPYYKLTWEKKKFKDEKQSLRASRLRAEMFAKQPVA 205  
 DB 404 EEEILRLHQKQKQSLSLRRLQKALIEEEMRMEEESQRLSWLRQVQSTQA-----457  
 QY 206 TQFLMDHDDEBDPLKTGLYSKRAAKSD-----DTSDDDFMEGSGEDGGS-----252  
 DB 458 -----DEDQIRAEQASLQKLEELSEQQAERASLQKXRMQLQKLEIEASEKSEQ 511  
 QY 253 -----DGMGDSGSEFLQDRFSETERY-----HTE-----SLQMSK 284  
 DB 512 AALNAKEXKALQQLREQLERKEAVATLEKHSALIELRLCSSLEAKHREVSSLQKIQ 571  
 QY 285 QELIKETLELEKCLSRMEDENNRLLESKRLGGDARVEL-----ELELDRL 332  
 DB 572 EAQKKEAQLOKQKLGQVE---HRVQKSYHVAQVEHLSLLEKQGEVGEHRRLDKM 628  
 QY 333 RAENLQILT---ENELHROQERAPL 354  
 DB 629 KEHQVMAKARQYEAERKORAEI 654  
 RESULT 11  
 Q95LST7 PRELIMINARY; PRT; 1455 AA.  
 AC Q95LST7 (TREMBlrel. 13, Created)  
 DT 01-MAY-2000 (TREMBlrel. 13, Last sequence update)  
 DT 01-OCT-2002 (TREMBlrel. 22, Last annotation update)  
 DE Hypothetical protein KIAA1052.  
 GN KIAA1052.  
 OS Homo sapiens (Human).  
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 OC Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.  
 OX NCBI\_TaxID=9606;  
 RN (1)  
 RP SEQUENCE FROM N.A.  
 RC TISSUE=Brain;  
 RX MEDLINE=99397452; PubMed=10470851;  
 RA Tanaka A., Kotani H., Nomura N., Ohara O.;  
 RT "Prediction of the coding sequences of unidentified human genes. XIV.  
 RT The complete sequences of 100 new cDNA clones from brain which code  
 RT for large proteins in vitro."  
 RL DNA Res. 6:197-205(1999).  
 DR EMBL: AB028975; BAA83004.1; -  
 DR InterPro: IPR001202; WW\_reps\_WWP.  
 DR Pfam: PF00397; WW; 1.  
 DR SMART: SMO0456; WW; 1.

DR PROSITE; PS50020; WW\_DOMAIN\_2; 1.  
 KM Hypothetical protein.  
 SQ SEQUENCE 1455 AA; 163543 MW; 7F48093100C34819 CRC64;  
 Query Match 7.7%; Score 147; DB 4; Length 1455;  
 Best Local Similarity 21.6%; Pred. No. 0.033;  
 Matches 97; Conservative 65; Mismatches 146; Indels 142; Gaps 19;  
 QY 25 QEELNRPFGAEEVP-----VPEEDSRWQSRF-----POLGPRGEGE-----64  
 DB 456 QDELQSKQSGLEERHRLSPPLPHE-RAQSPRLATEEPPQ-----GREGQPEWKEAEL 509  
 QY 65 -----GSLSEQPPPLQTA-CPESSCLREGEKQNGDSSAGDFFPPAE-----VE 110  
 DB 510 ADELGDSASLSTLQSLQREQAPSPPAACEKQKQHSQAELGPGQEAEDPEEKVAVS 569  
 QY 111 PTP-----EALLAQPCHDSEASKLGPAPAGSEEWQGOQROLGKK-----152  
 DB 570 PTPPVSPVSTEVAPPEQLSEAA-LKAMEAVAVQLEQDRHLLSKQKQKQQLREKL 628  
 QY 153 -----HRRRPSKKRHKMPYYKLTWEKKKFKDEKQSLRASRLRAEMFAKQPVA 201  
 DB 629 CQEEBEILRLHQKQKQSLSLRRLQKALIEEEMRMEEESQRLSWLRQVQSTQA--686  
 QY 202 PYNTTQFLMDHDDEBDPLKTGLYSKRAAKSD-----DTSDDDFMEGSGEDGGS- 252  
 DB 687 -----DEDQIRAEQASLQKLEELSEQQAERASLQKXRMQLQKLEIEASE 736  
 QY 253 -----DGMGDSGSEFLQDRFSETERY-----HTE-----SLQ 280  
 DB 737 KSEQALNAKEXKALQQLREQLERKEAVATLEKHSALIELRLCSSLEAKHREVSSLQ 796  
 QY 281 NMSKQELIKETLELEKCLSRMEDENNRLLESKRLGGDARVEL-----ELE 328  
 DB 797 KXIQEAQKQKQKQKQKLGQVE---HRVQKSYHVAQVEHLSLLEKQGEVGEHRR 853  
 QY 329 LQRLRAENLQILT---ENELHROQERAPL 354  
 DB 854 LDKMKHEHQVMAKARQYEAERKORAEI 883  
 RESULT 12  
 Q95LST7 PRELIMINARY; PRT; 560 AA.  
 AC Q95LST7 (TREMBlrel. 19, Created)  
 DT 01-DEC-2001 (TREMBlrel. 19, Last sequence update)  
 DT 01-OCT-2002 (TREMBlrel. 22, Last annotation update)  
 DE Hypothetical 63.1 kDa protein.  
 OS Macaca fascicularis (Crab eating macaque) (Cynomolgus monkey).  
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 OC Mammalia; Eutheria; Primates; Catarrhini; Cercopitheidae;  
 OC Cercopitheidae; Macaca.  
 OX NCBI\_TaxID=9541;  
 RN (1)  
 RP SEQUENCE FROM N.A.  
 RC TISSUE=Testis;  
 RA Hashimoto K., Osada N., Hida M., Kusuda J., Tanuma R., Hirai M.,  
 RA Terao K., Sugano S.;  
 RT "Isolation of novel full-length cDNA clones from macaque testis cDNA  
 RT libraries."  
 RL Submitted (SEP-2001) to the EMBL/GenBank/DBJ databases.  
 DR EMBL: AB071115; BAB64509.1; -  
 KM Hypothetical protein.  
 SQ SEQUENCE 560 AA; 63140 MW; C3EF0D9DC4635E36 CRC64;  
 Query Match 7.5%; Score 142.5; DB 6; Length 560;  
 Best Local Similarity 22.4%; Pred. No. 0.023;  
 Matches 88; Conservative 55; Mismatches 121; Indels 129; Gaps 19;  
 QY 27 ELNRPFGAEEVPEDSRWQSRFPPQ-----LGPRPGEKQSLSEQPPPLQ 75  
 DB 44 ESEPEEEEEESEEESEASQGTAADEQAKVPELTAAEAAGEEGP-GSPGRPARPOE- 101

```

QY 76 TCACPESSCLREGEKQNGDSDSAGDFFP-----PPAEVPTPEALLAOPCHDSEA- 127
DB 102 -----PB-----EPAAAGAEFPVQPKSGAGPELDAEAAAELEQAAEKEVYR 144
QY 128 -----SKLG-----APAA-----GGEEMGQOQOQUGKKKRR-----RSQKKRHHK 165
DB 145 SQASPLPTLRTIGEEBAAPAEATERVEEGEEDDEEERRRGAESEGRGAPKSGQEBCK 204
QY 166 P-----YKLTWEKKKKFDEKOSLRASRIAPMAFKGQVAPVYNTTQFLMDHDHDEBD 219
DB 205 PLGGRDEFEDLEKSE-----EVQKLOEQQRSDLLDQYR-----SLIMENRSQ----- 248
QY 220 LKTYGLY-----SKRAAASDDTSDDDFMEEGEEBDG-----SDMGDGSSEFLQDPF 267
DB 249 -RYNLVLQKRTPEALKKKKGLAEAVPDGAQAEAPKEKQAVYRLHGM-----LEDLKKQO 303
QY 268 SETTERHTESLONMSKQELIKYLEKCLSMEDENNRRLIESKRL-----G 316
DB 304 ADDLQWYH-----OEIGQLKKQCOEKLSTREVEKEMRRFQALKKQVYMWAMGSCRMKG 354
QY 317 GDDARVREL-----ELEDRRLAENLQ 339
DB 355 GRQALREVEQILALDEKKKEKMSAVRLLEVQL 387
RESULT 13
QY 09M475 PRELIMINARY; PRT; 1132 AA.
AC 09M475;
DT 01-MAY-2000 (TREMBlrel. 13, Created)
DT 01-OCT-2002 (TREMBlrel. 22, Last sequence update)
DT 01-MAR-2003 (TREMBlrel. 23, Last annotation update)
DE CG3108 protein.
GN CG3108.
OS Drosophila melanogaster (Fruit fly).
OC Eukaryota; Metazoa; Arthropoda; Hexapoda; Insecta; Pterygota;
OC Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;
OC Ephydroidea; Drosophilidae; Drosophila.
OC NCBI_TaxID=7227;
RN RP
RP SEQUENCE FROM N.A.
RC STRAIN=Berkley;
RX MEDLINE=20196006; PubMed=10731132;
RA Adams M.D., Celinker S.E., Holt R.A., Evans C.A., Gocayne J.D.,
RA Amaratunga P.G., Scherer S.E., Li P.W., Hoskins R.A., Galle R.F.,
RA George K.Y., Lewis S.E., Richards S., Ashburner M., Henderson S.N.,
RA Sutton G.G., Mortman J.R., Yandell M.D., Zhang Q., Chen L.X.,
RA Burton R.C., Rogers Y.-H.C., Blazef R.G., Champe M., Pfeiffer B.D.,
RA Wan K.H., Doyle C., Baxter E.G., Helt G., Nelson C.R., Miklos G.L.G.,
RA Abril J.F., Achapany A., An H.-J., Andrews-Pfambhoch C., Baldwin D.,
RA Ballew R.M., Basu A., Bakendale J., Bayraktaroglu L., Beasley E.M.,
RA Beeson K.Y., Benos P.V., Bertman B.P., Bhandari D., Bolshakov S.,
RA Borkova D., Botchan M.R., Bouck J., Brokstein P., Brotlier P.,
RA Burtis K.C., Busam D.A., Butler H., Cadieu E., Center A., Chandra I.,
RA Cherry J.M., Cawley S., Dahlke C., Davenport L.B., Davies P.,
RA de Pablo B., Delcher A., Deng Z., Mays A.D., Dew I., Dietz S.M.,
RA Dodeon K., Doup L.E., Domnes M., Dugan-Rocha S., Dunkov B.C., Dunn P.,
RA Durbin K.J., Evangelista C.C., Ferraz C., Fertler S., Fleischmann W.,
RA Foster C., Gagne J., Garg N.S., Gelbart W.M., Glasser K.,
RA Glodok A., Gong F., Gorrell J.H., Gu Z., Guan P., Harris M.,
RA Harris N.L., Harvey D., Heiman T.J., Hernandez J.R., Houck J.,
RA Hostin D., Houston K.A., Howland T.J., Wei M.-H., Idegam C.,
RA Jatali M., Kalush P., Kapten G.H., Ke Z., Kennison J.A., Ketchum K.A.,
RA Kimmel B.E., Kodira C.D., Kraft C., Kravitz S., Kulp D., Lai Z.,
RA Laske P., Lei Y., Levitsky A.A., Li J., Li Z., Liang Y., Lin X.,
RA Liu X., Mattei B., McIntosh T.C., McLeod M.P., McPherson D.,
RA Merkulov G., Mishina N.V., Mobarry C., Morris J., Moshrefi A.,
RA Mount S.M., Moy M., Murphy B., Murphy L., Muzny D.M., Nelson D.L.,
RA Nelson D.R., Nelson K.A., Nixon K., Nusser D.R., Pacleb J.M.,
RA Palazolo W., Pittman G.S., Pan S., Pollard J., Puri V., Reese M.G.,
RA Reinert K., Remington K., Saunders R.D.C., Scheeler F., Shen H.,
RA Shue B.C., Siden-Kiamos I., Simpson M., Skupski M.P., Smith T.,

```

```

RA Spier E., Spradling A.C., Stapleton M., Strong R., Sun E.,
RA Switskas R., Tector C., Turner R., Venter E., Wang A.H., Wang X.,
RA Wang Z.-Y., Wassarman D.A., Weinstock G.M., Weissbach J.,
RA Williams S.M., Woodage T., Worley K.C., Wu D., Yang S., Yao Q.A.,
RA Ye J., Yeh R.-F., Zaveri J.S., Zhan M., Zhang G., Zhao Q., Zheng L.,
RA Zheng X.H., Zhong F.N., Zhong W., Zhou X., Zhu S., Zhu X., Smith H.O.,
RA Gibbs R.A., Myers E.W., Rubin G.M., Venter J.C.;
RT "The genome sequence of Drosophila melanogaster."
RL Science 287:2185-2195 (2000).
RN [2]
RP SEQUENCE FROM N.A.
RA Celinker S.E., Adams M.D., Krommiller B., Wan K.H., Holt R.A.,
RA Evans C.A., Gocayne J.D., Amaratunga P.G., Brandon R.C., Rogers Y.,
RA Banson J., An H., Baldwin D., Banson J., Beeson K.Y., Busam D.A.,
RA Carlson J.W., Center A., Champe M., Davenport L.B., Dietz S.M.,
RA Dodson K., Dorsett V., Doup L.E., Doyle C., Drenek D., Farfan D.,
RA Fertler S., Frise E., Galle R.F., Garg N.S., George R.A.,
RA Gonzalez M., Houck J., Hoskins R.A., Hostin D., Howland T.J.,
RA Idegam C., Jatali M., Kruse D., Li P., Mattei B., Moshrefi A.,
RA McIntosh T.C., Moy M., Murphy B., Nelson C., Nelson K.A., Nunoo J.,
RA Pacleb J., Paragas V., Park S., Patel S., Pfeiffer B.,
RA Phouanavong S., Pittman G.S., Puri V., Richards S., Scheeler F.,
RA Stapleton M., Strong R., Switskas R., Tector C., Tyler D.,
RA Williams S.M., Zaveri J.S., Smith H.O., Venter J.C., Rubin G.M.;
RT "Sequencing of Drosophila melanogaster genome."
RL Submitted (MAR-2000) to the EMBL/GenBank/DBJ databases.
RN [3]
RP SEQUENCE FROM N.A.
RA Mistra S., Crosby M.A., Matthews B.B., Bayraktaroglu L., Campbell K.,
RA Hradecky P., Huang Y., Kaminker J.S., Prochnik S.E., Smith C.D.,
RA Tupy J.L., Bergman C., Bertman B., Carlson J.W., Celinker S.E.,
RA Clump M., Drysdale R., Emmert D., Frise E., de Grey A., Harris N.,
RA Krommiller B., Marshall B., Millburn G., Richter J., Russo S.,
RA Searle S.M.J., Smith E., Shu S., Smutnick F., Whitfield E.,
RA Ashburner M., Gelbart W.M., Rubin G.M., Mungall C.J., Lewis S.E.;
RT "Annotation of Drosophila melanogaster genome."
RL Submitted (MAR-2000) to the EMBL/GenBank/DBJ databases.
RN [4]
RP SEQUENCE FROM N.A.
RA Adams M.D., Celinker S.E., Gibbs R.A., Rubin G.M., Venter C.J.;
RL Submitted (MAR-2000) to the EMBL/GenBank/DBJ databases.
RN [5]
RP SEQUENCE FROM N.A.
RA Flybase;
RL Submitted (SEP-2002) to the EMBL/GenBank/DBJ databases.
DR EMBL: AE003435; AAF46083.2; -
DR HSSP: P00730; 2CTC.
DR Flybase; FBgn0029807; CG3108.
DR InterPro; IPR000834; Zn cardoSept.
DR Pfam; PF00246; Zn cardoSept; 1.
DR PRINTS; PR00765; CRBOXYPASA.
DR PROSITE; PS00132; CARBOXYPEPT_ZN_1; 1
SQ SEQUENCE 1132 AA; 125335 MW; 09B841A1C6D7E526 CRC64;
Query Match 7.4%; Score 141.5; DB 5; length 1132;
Best Local Similarity 22.3%; Pred. No. 0.062;
Matches 80; Conservative 57; Mismatches 138; Indels 83; Gaps 17;
QY 4 PLSLEYOHPOTSNTGAANOVEINLPPRPPCAEERVPEDSRWOSRAFPQUGRGPBG 63
DB 192 PLNDELPEDBEDEBPATTESAV--ELEKESEAMDOVPESESIQBPQVQ---GEYOS 246
QY 64 EG-----SLESOPPLQTOACPESSCLREGEKQNGDSDSAGDFFPAPAEVPTPEAE 116
DB 247 DEQAKTKEPTEIAQP---EVEAPPEAEAPAE-----PQLEVEPQPEVE 288
QY 117 LLAQCPHDSKASKLAPAPAGGEEGQOQOQUGKKKRRRPPSKKRRHMKVYKLTWEKK 176
DB 289 --SQPEVSQPEVAQPEVEPQSEVSQ-----PEASHSEPE---TQAEVE 330
QY 177 KDEKQSLRASIRAPMAFKGQVAPVYNTTQFLMDHDHDEBDLTKGLYSKRAAASDPT 236
DB 331 AQPESVSLPEASQPE--AESQP-----EREPEVE--AEKISDNEVD 369

```

QY 237 SDDDFMEEGEDGSDGDSGSEFLQDPSETERYHTESLQNNKQELIKYLELEK 236  
 DB 370 TEASLMETLIVE--GIEDGLTAAADNLVPEELAEASDKQETE--LESEDOQSPTVAIE--EQ 425  
 QY 297 CLSRMEPENRRLLESKRLGDDARVALELELDRLAENLQ--LTITENELHROQERAP 353  
 DB 426 AVPEIEQEKER---EPQITLAD-----ETEDSAPQSNPEEYELAEQHTAEAIAP 474

## RESULT 14

076153 PRELIMINARY; PRT; 443 AA.  
 AC 076153;  
 DT 01-NOV-1998 (TEMBLrel. 08, Created)  
 DT 01-NOV-1998 (TEMBLrel. 08, Last sequence update)  
 DT 01-OCT-2002 (TEMBLrel. 22, Last annotation update)  
 DE Rep60.  
 OS Periplaneta americana (American cockroach).  
 OC Eukaryota; Metazoa; Arthropoda; Hexapoda; Insecta; Pterygota;  
 OC Neoptera; Orthopteroidea; Dictyoptera; Blattaria; Blattellidae;  
 OC Blattellidae; Periplaneta.  
 NC NCB1\_TaxID=6978;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RA MEDLINE=20193551; PubMed=10727895;  
 RA Arai T., Kubo T., Natori S.;  
 RT "Identification, characterization and cDNA cloning of two novel  
 RT proteins secreted into the external space of the regenerating leg of  
 RT Periplaneta americana";  
 RL Insect Biochem. Mol. Biol. 30:287-295 (2000).  
 DR EMBL; AB012307; BAA32795.1;  
 SQ SEQUENCE 443 AA; 47987 MW; 91AF646B161C99B CRC64;

Query Match 7.4%; Score 140.5; DB 5; Length 443;  
 Best Local Similarity 20.4%; Pred. No. 0.024;

Matches 80; Conservative 66; Mismatches 140; Indels 107; Gaps 17;  
 QY 4 PFLSEYOHOPQNTCTGAAA-----VQELINPERPGAEERVPEDSRMOSRAFPOL 55  
 DB 47 PVAAPVPEKEPVAAATTVAAADVKEDEKPAEEVPPPAAPVVEEMKPAEDA----- 96  
 QY 56 GGRPGEGESLSQPPPLQTAQPE---SSCLREGEKGNGD-----DSSAGDPPPPA 107  
 DB 97 ---PSPAPEAPQPEVAV--PEAVPEVAQDETNNKGEVNDGSSDLKTEATSSQVAKPV 152  
 QY 108 E-----VEPTBEALIAQPCHDSESKLGAPAGGEEMGQOQOLGK 150  
 DB 153 EEKVALNAPVGEPAEKVPAEKKVEEVKVDQAAPTAAEPKAEKEEKPAKDE--K 209  
 QY 151 KKHRRRP-SKKKHHMKPYUKLTWEKKKPFDEKOSLRASRIRAEMFAKQPVAPYNTTQFL 209  
 DB 210 VEEAARPVSRKRTAP---KEEKKKPTKKA-----KAE----- 242  
 QY 210 MDDHDEEPDLTKGLYSKRAAASDDTSDDFMEEGEEDGSDGSDGSEFLQDPFSE 269  
 DB 243 -DVKVQEEFPQDVQVQVEEVKVDDEPKQBDVQVODAKKE-----EVKVEDAKE 291  
 QY 270 TYERYHTESLQNNKQELIKYLELEKLSRMEDENNRRLLESKRLGDD--DARVELE 326  
 DB 292 EEVAVQEEVQDAAKADVDV-----KVQDE---VLEEVKVESDAKEEVKEAK 341  
 QY 327 LELDRLEAENLQLTITENELHQ---QERAPLTK 356  
 DB 342 VEQDAKVQEDVKV--QDEVKQDAASAEVNPVSK 372

## RESULT 15

08TXA4 PRELIMINARY; PRT; 609 AA.  
 AC 08TXA4;  
 DT 01-JUN-2002 (TEMBLrel. 21, Created)  
 DT 01-JUN-2002 (TEMBLrel. 21, Last sequence update)

DT 01-MAR-2003 (TEMBLrel. 23, Last annotation update)  
 DE Uncharacterized protein.  
 GN MK0771.  
 OS Methanopyrus kandleri.  
 OC Archaea; Euryarchaeota; Methanopyri; Methanopyrales; Methanopyraceae;  
 OC Methanopyrus.  
 NC NCB1\_TaxID=2320;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=AV19 / DSM 6324 / JCM 9639;  
 RX MEDLINE=21927647; PubMed=11930014;  
 RA Stearney A.I., Mezhevaya K.V., Makarova K.S., Polunin N.N.,  
 RA Sncherbina O.V., Shakhova V.V., Belova G.I., Aravind L.,  
 RA Naitale D.A., Rogozin I.B., Tatunov R.L., Wolf Y.I., Shtetler K.O.,  
 RA Malayh A.G., Koonin E.V., Kozavkin S.A.;  
 RT "The complete genome of hyperthermophilic Methanopyrus kandleri AV19  
 RT and monophyly of archaeal methanogens";  
 RL Proc. Natl. Acad. Sci. U.S.A. 99:4644-4649 (2002).  
 DR EMBL; AE010369; AAM01985.1;  
 DR HSSP; P04268; IIC2.  
 DR InterPro; IPR002017; Spectrin.  
 KM Complete proteome.  
 SQ SEQUENCE 609 AA; 69552 MW; AB10C9780DC5AD78 CRC64;

Query Match 7.3%; Score 140; DB 17; Length 609;  
 Best Local Similarity 22.0%; Pred. No. 0.039;

Matches 80; Conservative 66; Mismatches 147; Indels 70; Gaps 17;  
 QY 31 ERPPGAER-----VEEDSRMOSRAFPOLGRRPGEGSGS-----LSQPPPLQTO 77  
 DB 5 KKKGGRRERPEIDLKIPPEBEGAPKTLPPBEGPG--ESGAELAVPLKLPKPP----- 58  
 QY 78 ACPESSCLREGEKGNGDSSAGDPFPPAEVE--PTPEALLAOPCHDSEASKLGAPAG 136  
 DB 59 -----KPSKKPPEGESEKKKKERRPPALIKPPRPREERTAPSVLEAELRLKAKENK 110  
 QY 137 GEE--EW-----GQOQROLGK-KKHRRRPSKKRHHKPYUKLTWEKKKPFDE---- 180  
 DB 111 LREELDEMNKAKSAMGERDRLRSEIKRLKELEKQEKELDYIKISKQLEKLEKAKRE 170  
 QY 181 ----KQSLRASRIRAEMFAKQPVAPYNTTQFLMDHDEEDDLTKGLYSKRAAKSDDT 236  
 DB 171 SEELKKAERYERYEKIA-----GKYNELKSLLELSDONRLLENL--KTLKEKYNEI 223  
 QY 237 SDD-DFMEEGEEDGSDGSDGSEFLQDPSET--YER-----YHTESLQNNK---KQEL 287  
 DB 224 KEERDLKETETVEGKTK---DQAKOSKLYKESERDDLANEVEALRNENKLRKKI 279  
 QY 288 IKYLELEKLSRMEDENNRRLLESKRLGDDARVALELELDRLAENLQLTITENELH 347  
 DB 280 DKLKSELNLIQKTLKDRKKLEKAKOHIGKLRREIRKDEEIRKLKAKOSKL--KDEIKR 337  
 QY 348 QOE 350  
 DB 338 YEE 340

Search completed: February 5, 2004, 13:31:02  
 Job time : 112 secs

GenCore version 5.1.6  
Copyright (c) 1993 - 2004 Compugen Ltd.

OM protein - nucleic search, using frame\_plus\_p2n model

Run on: February 5, 2004, 13:38:50 / Search time 2658 Seconds  
(without alignments)  
3282.659 Million cell updates/sec

Title: US-09-972-758a-2  
Perfect score: 1910  
Sequence: 1 MAEPRISEXQHPQTSNCTG.....LTENELHROERAPLSKFGD 359

Scoring table: BLOSUM62  
Xgapop 10.0, Xgapext 0.5  
Ygapop 10.0, Ygapext 0.5  
Fgapop 6.0, Fgapext 7.0  
Delop 6.0, Delext 7.0

Searched: 22781392 seqs, 12152238056 residues

Total number of hits satisfying chosen parameters: 45562784

Minimum DB seq length: 0  
Maximum DB seq length: 200000000

Post-processing: Minimum Match 0%  
Maximum Match 100%  
Listing first 45 summaries

Command line parameters:  
-MODE=frame\_plus\_p2n.model -DEV=xlh  
-Q=/cgn2\_1/USPTO.spool/US09972758/runat\_05022004\_095003\_1203/app.query.fasta.1.519  
-DB=EST -QFMT=fastap -SUFFIX=xt -MINMATCH=0.1 -LOOPEL=0 -LOOPEXT=0  
-UNITS=bits -START=1 -END=1 -MATRIX=blonum62 -TRANS=human40.cdi -LIST=45  
-DOCALIGN=200 -THR SCORE=pct -THR MAX=100 -THR MIN=0 -ALIGN=15 -MODE=LOCAL  
-OUTFMT=ptc -NOM=ext -HEAPSIZE=500 -MINLEN=0 -MAXLEN=200000000  
-USER=US09972758.@CNCN\_1.1.2810@runat\_05022004\_095003\_1203 -NCPU=6 -ICPU=3  
-NO\_MMAP -LARGEQUERY -NEG SCORES=0 -WAIT -DSBLOCK=100 -LONGCLOC  
-DEV TIMEOUT=120 -WARN TIMEOUT=30 -THREADS=1 -XGAPOP=10 -XGAPEXT=0.5 -FGAPOP=6  
-FGAPEXT=7 -YGAPOP=10 -YGAPEXT=0.5 -DELOP=6 -DELEXT=7

Database : EST.\*  
1: em\_estba.\*  
2: em\_esthum.\*  
3: em\_estin.\*  
4: em\_estmu.\*  
5: em\_estrov.\*  
6: em\_estropl.\*  
7: em\_estro.\*  
8: em\_hic.\*  
9: gb\_est1.\*  
10: gb\_est2.\*  
11: gb\_hic.\*  
12: gb\_est3.\*  
13: gb\_est4.\*  
14: gb\_est5.\*  
15: em\_estfun.\*  
16: em\_estcom.\*  
17: em\_ges\_hum.\*  
18: em\_ges\_inv.\*  
19: em\_ges\_pin.\*  
20: em\_ges\_vrt.\*  
21: em\_ges\_fun.\*  
22: em\_ges\_mam.\*  
23: em\_ges\_mus.\*  
24: em\_ges\_pro.\*  
25: em\_ges\_rod.\*  
26: em\_ges\_phg.\*  
27: em\_ges\_vrl.\*  
28: gb\_ges1.\*

29: gb\_ges2.\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

# SUMMARIES

Result No.	Score	Query	Match Length	DB ID	Description
1	1414.5	74.1	894	14	CD518239
2	1350	70.7	1201	9	AL552998
3	1327.5	69.5	903	13	BQ221641
4	1307	68.4	750	13	BX112898
5	1266.5	66.3	771	12	BI871190
6	1231	64.5	797	12	BF984049
7	1163.5	60.9	797	12	BF819109
8	1142.5	59.8	832	13	BU170664
9	1137.5	59.6	830	13	BU526446
10	1129	59.1	750	9	AV726891
11	1120	58.6	852	12	BI091005
12	1120	58.6	1078	9	AL578242
13	1111.5	58.2	871	10	BG291150
14	1074	56.2	1294	12	BM474968
15	1059	55.4	683	14	CA430432
16	1056.5	55.3	1080	12	BM474949
17	1042	54.6	949	10	BG165450
18	996	52.1	1201	9	AL576112
19	963	50.4	557	12	BM129736
20	953	49.9	762	10	BF167480
21	946.5	49.6	745	10	BS547281
22	933	48.8	907	10	BF184535
23	932.5	48.8	666	12	BG864490
24	909	47.6	810	10	BF607249
25	905	47.4	999	12	BI141874
26	900	47.1	933	10	BF674587
27	893	46.8	535	14	CA865829
28	879	46.0	563	9	AM65096
29	879	46.0	843	10	BR026064
30	868.5	45.5	807	13	BU461116
31	860	45.0	638	9	AA546538
32	858	44.9	591	13	BU471258
33	857.5	44.9	765	13	BU489324
34	857.5	44.9	727	13	BU400268
35	852.5	44.6	920	13	BU249231
36	851	44.6	547	9	AT796944
37	850	44.5	410	9	AL555845
38	841	44.0	482	9	AI948609
39	841	44.0	566	10	AW962738
40	829.5	43.4	871	13	BU314320
41	826.5	43.3	656	13	BU287394
42	826.5	43.3	911	13	BU288715
43	822	43.0	462	9	AV728223
44	814.5	42.6	741	13	BU374772
45	812	42.5	907	12	BI412550

## ALIGNMENTS

RESULT 1  
LOCUS CD518239 894 bp mRNA linear EST 06-JUN-2003  
DEFINITION AGENCOURT\_14375843 NIH\_MGC\_181 Homo sapiens cDNA clone  
IMAGE:30407314 5', mRNA sequence.  
ACCESSION CD518239  
VERSION CD518239.1 GI:31449957  
KEYWORDS EST.  
SOURCE Homo sapiens (human)  
ORGANISM Homo sapiens  
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.  
REFERENCE 1 (bases 1 to 894)



BASE COUNT 295 a 280 c 360 g 200 t 66 others  
ORIGIN

## Alignment Scores:

Pred. No.: 1,88e-70 Length: 1201  
Score: 1350.00 Matches: 262  
Percent Similarity: 95.99% Conservative: 1  
Best Local Similarity: 95.62% Mismatches: 10  
Query Match: 70.68% Indels: 4  
Gaps: 1

US-09-972-758a-2 (1-359) x AL552998 (1-1201)

QY 1 MetAlaGluProPheLeuSerGluTyrGlnHisGlnProGlnThrSerAsnCysThrGly 20  
Db 202 ATGGCGGAGCCATCTTGTTCAGAAATATCAACCAAGCTAACTGACAACTGACAGK 261  
QY 21 AAlaAlaValGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGln 40  
Db 262 GCTGCTGCTGCTCAG 321  
QY 41 ProGlnGluAspSerArgTrpGlnSerArgAlaPheProGlnLeuGlyGlyArgProGly 60  
Db 322 CCGAG 381  
QY 61 ProGlnGlyGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGln 80  
Db 382 CCGAG 441  
QY 81 GluSerSerCysLeuArgGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGln 100  
Db 442 GAATCTAGCTGCTGAG 501  
QY 101 GlyAspPheProProProAlaGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGln 120  
Db 502 GGGGAGCTTCCCGGCGGCGGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 561  
QY 121 ProCysHisAspSerGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGln 140  
Db 562 CTTTTCATGACTCCAG 620  
QY 141 TrpGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGln 160  
Db 621 TGGGAGCAG 680  
QY 161 LysArgHisTrpLysProTyrTyrLysLeuThrTrpGlnGlnGlnGlnGlnGlnGln 180  
Db 681 AAGCGGCACTTGGAAACCGTACTCAAGCTGACCTGGGAGAGAGAGAGAGAGAGAGAG 740  
QY 181 LysGlnSerLeuArgAlaSerArgAlaGlnGlnGlnGlnGlnGlnGlnGlnGlnGln 200  
Db 741 AAACAG 800  
QY 201 AlaProTyrAsnThrThrGlnPheLeuMetAspAspHisAspGlnGlnGlnGlnGln 220  
Db 801 GCGCCCTATTAACACACAG 860  
QY 221 LysThrGlyLeuTyrSerLysArgAlaAlaLysSerAspAspThrSerAspAspAsp 240  
Db 861 AAAACCGGAG 920  
QY 241 PheMetGluGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGln 260  
Db 921 TTATGATGAAG 979  
QY 261 GluPheLeuGlnArgAspPheSerGluThrTyrGlnGlnGlnGlnGlnGlnGlnGln 274  
Db 980 AGTTT---TSMGCGGAG 1017

RESULT 3  
BO221641 903 bp mRNA linear EST 02-MAY-2002  
LOCUS BO221641  
DEFINITION AGENCOURT\_7552882 NIH\_MGC\_68 Homo sapiens cDNA clone IMAGE:6058110

ACCESSION 5', mRNA sequence.  
VERSION BO221641  
KEYWORDS BO221641.1 GI:20403041  
SOURCE EST.  
ORGANISM Homo sapiens (human)

REFERENCE Homo sapiens  
AUTHORS Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
TITLE Mammalia; Euteleostomi; Primates; Catarrhini; Homidae; Homo.  
JOURNAL 1 (bases 1 to 903)  
COMMENT NIH-MGC http://mgs.nci.nih.gov/

CONTACT: Robert Strausberg, Ph.D.  
Email: cga@bbs-rcmail.nih.gov  
Tissue Procurement: DCTD/DTF/Gazdar  
CDNA Library Preparation: Life Technologies, Inc.  
CDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)  
DNA Sequencing by: Agencourt Bioscience Corporation  
Clone distribution: MGC clone distribution information can be  
found through the I.M.A.G.E. Consortium/LLNL at:  
http://image.llnl.gov  
Plate: LLNL1332 row: K column: 07  
High quality sequence stop: 644.  
Location/Qualifiers

## FEATURES

## source

1..903  
/organism="Homo sapiens"  
/mol\_type="mRNA"  
/db\_xref="taxon:9606"  
/clone="IMAGE:6058110"  
/tissue\_type="large cell carcinoma"  
/lab\_host="DH10B (phage-resistant)"  
/clone\_idb="NIH MGC 68"  
/note="Organ: lung; Vector: pCMV-Sport; Site 1: NotI;  
Site 2: SalI; Cloned unidirectionally. Primer: Oligo dT.  
Average insert size 1.8 kb. Library constructed by Life  
Technologies."

BASE COUNT 217 a 259 c 309 g 115 t 3 others  
ORIGIN

Alignment Scores:  
Pred. No.: 3.15e-69 Length: 903  
Score: 1327.50 Matches: 251  
Percent Similarity: 97.68% Conservative: 2  
Best Local Similarity: 96.91% Mismatches: 4  
Query Match: 69.50% Indels: 2  
Gaps: 1

US-09-972-758a-2 (1-359) x BO221641 (1-903)

QY 51 AlaPheProGlnLeuGlyGlyArgProGlnGlnGlnGlnGlnGlnGlnGlnGlnGln 70  
Db 1 GCGTTCCTCCCAAGTTGGGTGGCTCCGGGCGGAGAGAGAGAGAGAGAGAGAGAGAGAG 60  
QY 71 ProProProLeuGlnThrGlnAlaCysProGlnSerSerCysLeuArgGlnGlnGln 90  
Db 61 CCACTCCCTTCAG 120  
QY 91 GlyGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGln 110  
Db 121 GCGCAGAAAGAA 180  
QY 111 ProThrProGlnAlaGlnLeuLeuAlaGlnProCysHisAspSerGlnLysSerLys 130  
Db 181 CCGAGCCCGGAG 240  
QY 131 GlyAlaProAlaAlaGlyGlyGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGln 150  
Db 241 GGGGCTCTGCGGAG 300  
QY 151 LysLysHisArgArgArgProSerLysLysLysLysLysLysLysLysLysLysLys 170  
Db 301 AAAAATCATGAG 360







RESULT 5  
BI871190 771 bp mRNA linear EST 11-OCT-2001  
LOCUS 603395081F1 NIH\_MGC\_90 Homo sapiens CDNA clone IMAGE:5404724 5',  
DEFINITION mRNA sequence.  
ACCESSION BI871190.1 GI:16044865  
VERSION BI871190.1  
KEYWORDS EST.  
SOURCE Homo sapiens (human)  
ORGANISM Homo sapiens  
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
Mammalia; Eutheria; Primates; Catarrhini; Homiidae; Homo.  
REFERENCE 1 (bases 1 to 771)  
NIH-MGC http://mhc.nci.nih.gov/  
TITLE National Institutes of Health, Mammalian Gene Collection (MGC)  
JOURNAL Unpublished  
COMMENT Contact: Robert Strausberg, Ph.D.  
Email: cgapbs-remail.nih.gov  
Tissue Procurement: ATCC  
CDNA Library Preparation: Life Technologies, Inc.  
DNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)  
DNA Sequencing by: Incyte Genomics, Inc.  
Clone distribution: MGC clone distribution information can be  
found through the I.M.A.G.E. Consortium/LLNL at:  
http://image.llnl.gov  
Plate: LLAM12032 row: b column: 21  
High quality sequence stop: 733.  
Location/Qualifiers

FEATURES  
source  
1..771  
/organism="Homo sapiens"  
/mol\_type="mRNA"  
/db\_xref="taxon:9606"  
/clone="IMAGE:5404724"  
/tissue\_type="adenocarcinoma, cell line"  
/lab\_host="DH10B (phage-resistant)"  
/note="Organ: liver; Vector: pCMV-SPORT6; Site 1: NotI;  
Site 2: SalI; Cloned unidirectionally; oligo-dT primed.  
Average insert size 1.7 kb. Library enriched for  
full-length clones and constructed by Life technologies.  
Note: this is a NIH-MGC Library."

BASE COUNT 187 a 213 c 273 g 98 t  
ORIGIN

Alignment Scores:  
Pred. No.: 1.09e-65 Length: 771  
Score: 1266.50 Matches: 248  
Percent Similarity: 96.51% Conservative: 1  
Best Local Similarity: 96.12% Mismatches: 5  
Query Match: 66.31% Indels: 5  
DB: 12 Gaps: 1

US-09-972-758a-2 (1-359) x BI871190 (1-771)

QY 21 AIAAIAAIAValGluGluGluLeuAAspProGluValAlaGluGluAVal 40  
DB 3 GGTGCTGCTGCTCAAGAAAGCTGAACCTTGAAGCCGCCCGAGGAGAGGAGG 62  
QY 41 ProGluGluAAspSerArgTrrpGlnSerArgAlaPheProGluLeuGlyValArgProGly 60  
DB 63 CCGGAGAGAGACAGTGTGGCAATGAGAGGCTTCCCAAGTGGGTGGCGCTCGGGG 122  
QY 61 ProGluGlyValGlySerLeuGlnProGluProGluGlnThrGlnAlaCysPro 80  
DB 123 CCGAGAGGGGAGAGGAGCTGGAATCCCAACCCCTTGGAGCCAGGCGCTGTCCA 182  
QY 81 GluSerSerCysLeuArgGluGlyGlyValGlnAspGlyAAspAAspSerSerAlaGly 100  
DB 183 GAATCTAGCTGCTGAGAGAGGCGGAGAGGCGGAGAGAGAGAGAGAGAGAGAG 242  
QY 101 G1AAspPheProProAlaGluValGluProThrProGluValAlaGluLeuLeuAlaGln 120  
DB 243 GGCAGACTCCCGCGCGGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 302

QY 121 ProCysHisAspSerGluValAspLeuGluValAlaProAlaAlaGlyValGluGluGlu 140  
DB 303 CTTTGTATGACTCCGAGAGCGAGTAAAGTTGGGGGCTCTCTCGCGGAGGGGCGAAGAGAG 362  
QY 141 TrpGlyGlnGlnGlnArgGlnLeuGlyValValValValValValValValValValVal 160  
DB 363 TGGGAGACAGCAGCAG 422  
QY 161 LysArgHisTrpLysProTyrTyrLysLeuThrTrpGluGluValLysLysPheAspGlu 180  
DB 423 AAGCGGCAATGGAAACCGTACTACAGCTGACCTGGAGAGAGAGAGAGAGAGAGAGAGAG 482  
QY 181 LysGlnSerLeuArgAlaSerArgLysArgAlaGluMetPheAlaLysGlyValProVal 200  
DB 483 AAGAGAGCTTCCGAGCTTCAAGATCCGAGCGGAGAGAGAGAGAGAGAGAGAGAGAGAG 542  
QY 201 AlaProTyrAsnThrThrGlnPheLeuMetAspAspHisAspGlnGluGluProAspLeu 220  
DB 543 GCGCCCTTAAACACACAGCAGATTCCTCATGATGATCAAGACAGAGAGAGAGAGAGATCTC 602  
QY 221 LysThrGlyLeuTyrSerLysArgAlaAlaLysSerAspAspThrSerAspAsp 240  
DB 603 AAAACCGGCTGTACTCCAGAGCGGCGCCGCGGCAATCCGAGAGAGAGAGAGAGAGAGAG 661  
QY 241 PheMetGluGluGlyValGluGluLeuAspGlyLysSerAspGlyMetGlyValValValVal 259  
DB 662 TTCAATGAG 721  
QY 260 SerGlu-PheLeuGlnArgAsp-PheSerGlu-ThrTyrGluArgTyr 274  
DB 722 AGCGAGGTTCTGCGCAGCGGAACTTCTCGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 769

RESULT 6  
BF984049 982 bp mRNA linear EST 23-JAN-2001  
LOCUS 602306989P1 NIH\_MGC\_88 Homo sapiens CDNA clone IMAGE:4398238 5',  
DEFINITION mRNA sequence.  
ACCESSION BF984049.1 GI:12386861  
VERSION BF984049.1  
KEYWORDS EST.  
SOURCE Homo sapiens (human)  
ORGANISM Homo sapiens  
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
Mammalia; Eutheria; Primates; Catarrhini; Homiidae; Homo.  
REFERENCE 1 (bases 1 to 982)  
NIH-MGC http://mhc.nci.nih.gov/  
TITLE National Institutes of Health, Mammalian Gene Collection (MGC)  
JOURNAL Unpublished  
COMMENT Contact: Robert Strausberg, Ph.D.  
Email: cgapbs-remail.nih.gov  
Tissue Procurement: ATCC  
CDNA Library Preparation: Life Technologies, Inc.  
DNA Sequencing by: Incyte Genomics, Inc.  
Clone distribution: MGC clone distribution information can be  
found through the I.M.A.G.E. Consortium/LLNL at:  
http://image.llnl.gov  
Plate: LLAM10100 row: a column: 23  
High quality sequence stop: 691.  
Location/Qualifiers

FEATURES  
source  
1..982  
/organism="Homo sapiens"  
/mol\_type="mRNA"  
/db\_xref="taxon:9606"  
/clone="IMAGE:4398238"  
/tissue\_type="quodanal adenocarcinoma, cell line"  
/lab\_host="DH10B (phage-resistant)"  
/clone\_lib="NIH-MGC 88"  
/note="Organ: small intestine; Vector: pCMV-SPORT6;  
Site 1: NotI; Site 2: SalI; Cloned unidirectionally;  
oligo-dT primed. Average insert size 1.767 kb. Library  
enriched for full-length clones and constructed by Life

BASE COUNT 261 a 235 c 320 g 166 t  
 ORIGIN Technology: Note: this is a NIH\_MGC Library."

## Alignment Scores:

Pred. No.:	1 66e-63	Length:	982
Score:	1231.00	Matches:	238
Percent Similarity:	99.17%	Conservative:	0
Best Local Similarity:	99.17%	Mismatches:	2
Query Match:	64.45%	Indels:	1
DB:	10	Gaps:	0

US-09-972-758a-2 (1-359) x BF984049 (1-982)

```

QY 120 GlnProCyH1aAsePseGua1aSerLybLeuGlyAlaProAla1aGlyGlyGlu 139
DB 8 CAGCCTTGATCATGCTCCGAGGCCAGTAAAGTTGGGGCTCTCCCGCAGGGGCGAAG 67
QY 140 GltTtGlyGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGln 159
DB 68 GAGTGGGACAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAG 127
QY 160 LylbLysAlaH1aTtPsePsePsePsePsePsePsePsePsePsePsePsePse 179
DB 128 AAGAACCGGCTTGGAAACCGTCTCAAGCTGAGCTGGAGAGAGAGAGAGAGAGAG 187
QY 180 GltLysGlnSerLeuA1aSerArg1aLeuA1aGlnPheA1aLysGlyGlnPro 199
DB 188 GAGAAACAGAGCCTTGAGCTTCAAGATCCGAGCCGAGATGTTCCCAAGGCCAGCCG 247
QY 200 ValAlaProTyrAsnThrThrGlnPheLeuMetAspAspHisAspGlnGluPro 219
DB 248 GTCCGCCCTATTAACCAACGACGCTCCATGATGATCAACACAGAGAGAGCCGAT 307
QY 220 LeuLysThrGlyLeuTyrSerLysArgAlaAlaAlaLysSerAspAspThrSerAsp 239
DB 308 CTCAAACCGGCTGATCTCAAGCCGAGCCGCGCCAAATCCACAGACACACAGCTGAC 367
QY 240 AspPheMetGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGln 259
DB 368 GACTTATGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 427
QY 260 SerGluPheLeuGlnArgAspPheSerGluThrTyrGluArgTyrHisThrGluSerLeu 279
DB 428 AGGAGATTTCTGAGCGGAGCTTCTCGAGACCTAAGAGCGGATACCAACGAGACCTCG 487
QY 280 GlnAspMetSerLysGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGln 299
DB 488 CAGAACATGAGCAGAGAGAGCTCATCAAGAGATCTGAACTGAGAGAGAGAGAGAG 547
QY 300 ArgMetGluAspGluAsnAsnArgLeuArgLeuGlnSerLysArgLeuGlnGlnGln 319
DB 548 CCGATGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 607
QY 320 AlaArgValArgGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGln 339
DB 608 GCGCGGTGCGGAGCTGAGAGCTGAGAGCTGAGAGCTGAGAGCTGAGAGCTGAG 667
QY 340 LeuThrGlnAsnGlnLeuHisArgGlnGlnGlnGlnGlnGlnGlnGlnGlnGln 359
DB 668 CTGACCGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 726

```

RESULT 7  
 LOCUS BG819109 797 bp mRNA linear EST 22-MAY-2001  
 DEFINITION 602781202F1 NCI\_CGAP\_Brn67 Homo sapiens CDNA clone IMAGE:4932025  
 5', mRNA sequence.

ACCESSION BG819109  
 VERSION BG819109.1 GI:14166696  
 KEYWORDS EST.  
 SOURCE Homo sapiens (human)  
 ORGANISM Homo sapiens  
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;

REFERENCE  
 1 (bases 1 to 797)  
 NIH-MGC <http://mgi.nci.nih.gov/>.  
 AUTHORS National Institutes of Health, Mammalian Gene Collection (MGC)  
 JOURNAL Unpublished  
 COMMENT Contact: Robert Strausberg, Ph.D.  
 Email: [cgabbs-rc@mail.nih.gov](mailto:cgabbs-rc@mail.nih.gov)  
 Tissue Procurement: David N. Louis, M.D.  
 CDNA Library Preparation: Life Technologies, Inc.  
 DNA Sequencing by: Incyte Genomics, Inc.  
 Clone distribution: MGC clone distribution information can be found through the I.M.A.G.E. Consortium/LLNL at:  
<http://image.llnl.gov>  
 Plate: LLNL10857 row: c column: 02  
 High quality sequence stop: 797.  
 Location/Qualifiers

FEATURES  
 source 1..797  
 /organism="Homo sapiens"  
 /mol\_type="mRNA"  
 /db\_xref="taxon:9606"  
 /clone="IMAGE:4932025"  
 /tissue\_type="anaplastic oligodendroglioma with 1p/19q loss"  
 /lab\_host="DH10B (T1 phage-resistant)"  
 /clone\_11b="NCI\_CGAP\_Brn67"  
 /note="Organ: brain; Vector: PCMV-SPORT6; Site 1: NotI; Site 2: SalI; Cloned unidirectionally. Primer: Oligo dt. Average insert size 2.3 kb. Constructed by Life Technologies. Note: this is a NCI\_CGAP Library."

BASE COUNT 207 a 223 c 248 g 119 t  
 ORIGIN

## Alignment Scores:

Pred. No.:	1 34e-59	Length:	797
Score:	1163.50 <td>Matches:</td> <td>232 </td>	Matches:	232
Percent Similarity:	95.12% <td>Conservative:</td> <td>2 </td>	Conservative:	2
Best Local Similarity:	94.31% <td>Mismatches:</td> <td>8 </td>	Mismatches:	8
Query Match:	60.92% <td>Indels:</td> <td>6 </td>	Indels:	6
DB:	12 <td>Gaps:</td> <td>1 </td>	Gaps:	1

US-09-972-758a-2 (1-359) x BG819109 (1-797)

```

QY 1 MetAlaGluProPheLeuSerGluTyrGlnHisGlnProGlnThrSerAsnCyThrGly 20
DB 69 ATGGCCGAGCCTATCTTGCAGATATCAACACAGCCTCAAACTAGCAACTGTACAGGT 128
QY 21 AlaAlaAlaValGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGln 40
DB 129 GCTGCTGCTGCTCAGAGAGAGCTGAACCTGAGCCGCCCGGCGGAGAGAGCGGAG 188
QY 41 ProGlnGluAspSerArgTyrGlnSerArgAlaPheProGlnLeuGlyGlyArgProGly 60
DB 189 CCGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 248
QY 61 ProGlnGlyGlnGlySerLeuGlnSerGlnProProProProLeuGlnThrGlnAlaCyPro 80
DB 249 CCGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 308
QY 81 GluSerSerCysLeuArgGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGln 100
DB 309 GAATCTAGCTGCTTGAAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 368
QY 101 GlysAspPheProProProAlaGluValGluProThrProGlnAlaGluLeuLeuAlaGln 120
DB 369 GCGAGCTTCCCGCCCGCCGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 428
QY 121 ProCyHisAspSerGlnAlaSerLysLeuGlnGlnGlnGlnGlnGlnGlnGlnGlnGln 140
DB 429 CTTGTGATGATCTCCGAGCCAGTAAAGTTGGGGCTCTCCGCCGAGGGGCGAAGAGAG 488
QY 141 TtGlyGlnGlnGlnArgGlnLeuGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGln 160

```





	QY	14	TtPglvIngnInglnIuaVxgUlnLeuglYlyvsyIsyShIsAgaGAGzProSeRysLyS	160
	Db	122	TGGGACAGCGCAGAGACAGACTGGGAGNANNAACCTTAGAGACCCTTCAGAG	181
	QY	161	LysArgHisTrpLysProTyrrTyrLysLeuThrTrpGluJulysLysPheAspGlu	180
	Db	182	AAGCGCATTTGGAAACCGTATCTAACAGCTGAACCTGGAGNAGAAG--	226
	QY	181	LysGlnSerLeuAlaGlaSerArgIleArgAlaGluMetPheIleAlaLysGlnProVal	200
	Db	227	-----ATCGCCAAAGGGCCAGCCGTGC	247
	QY	201	AlaProTYraenThrThrglnPheLeuMetAspAspHisasapGlnGluGluProAspleu	220
	Db	248	GCGCCCTRTAACACCAACGCAAGTCTCTCATGTAGATATCACACACAGAGAGACCGGACTC	307
	QY	221	LysThrglyLeuTYrSerLYeSerAlaArgAlaAlaLysSerAspAspThrSerAspAsp	240
	Db	308	ANAAACCGGCCCTGTACTCCAAAGCGGGCCGNCGCCAAATCCAGACACACAGCGATGACGAC	367
	QY	241	PheMetGluGluGlyGlyGlnGluAspGlyLysSerAspGlymetGlyLysAspGlySer	260
	Db	368	TTTCATGTGAAGAGGGGGTGAGAGAGATGAGNGACGACGATGGATGTGGAGGGGAGACGGCAGC	427
	QY	261	GluPheLeuGlnArGaaspPheSerGluThrTYrGluArGYrHIsThrgluserLeugln	280
	Db	428	GAGTTTCTGCAGCGGGGACTTCTCCGAGACGTRACAGCGGTACACACAGGAGACCTGCAG	487
	QY	281	AsnMetSerLYSGlnGluLeuIleLysGlnTYrLeuglnLeuGlyLysCyLeusSerArg	300
	Db	488	AACATGTAGCAACGACGAGACTCATCAAGAGATGACTCGTAACCTGGAGNATGGCTTTCGCGC	547
	QY	301	MetGluAspGluAsnAsnArgLeuArgLeuglnserLYsaArgLeuGlyLysAspAla	320
	Db	548	ATGAGAGCGAGAAACACCGCTGCGCTGGAGAGACAAGCGCTGGGTGGAGACGACGCG	607
	QY	321	ArgValArgGluLeuGlnLeuGlnLeuLysAspArgLeuAlaGluAsnLeuGlnLeuLeu	340
	Db	608	CCTGTGCGGAGACTTGGAGCTCGAGCTTATACCCGCTTANNCCGCGAGACCTTCACTGCTG	667
	QY	341	ThrGluAsnGluLeuHisArgGlnGlnGluArgAlaProLeuSerLYs	356
	Db	668	ACCGANAAACGAACCTGACCCCGACAGCANAGCGAGCGCGCTTCCAAAG-715	
RESULT_11				
BIO91005			852 bp	mRNA linear EST 20-JUN-2001
LOCUS				
DEFINITION			60285366791 NIH_MGC_10 Homo sapiens cDNA clone IMAGE:4995065 5',	
ACCESSION			mRNA sequence.	
VERSION			BIO91005	
KEYWORDS			BIO91005.1 GI:14509335	
SOURCE			EST.	
ORGANISM			Homo sapiens (human)	
			Homo sapiens	
			Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;	
			Mammalia; Eutheria; Primates; Catarrhini; Hominiidae; Homo.	
REFERENCE			NIH-MGC http://mgs.nci.nih.gov/	
AUTHORS			National Institutes of Health, Mammalian Gene Collection (MGC)	
TITLE			Unpublished	
JOURNAL			Contact: Robert Strausberg, Ph.D.	
COMMENT			Email: cgabbs-remail.nih.gov	
			Tissue Procurement: ATCC	
			cDNA Library Preparation: Life Technologies, Inc.	
			cDNA Library Arrayed by: Incyte Genomics, Inc.	
			Clone Sequencing by: Incyte Genomics, Inc.	
			Data distribution: MGC clone distribution information can be	
			found through the I.M.A.G.E. Consortium/LNLN at:	
			http://image.llnl.gov	
			Plate: LLM11018 row: e column: 18	
			High quality sequence stop: 700.	
FEATURES			Location/Qualifiers	

	source	1..852	/organism="Homo sapiens"
			/mol_type="mRNA"
			/db_xref="taxon:9606"
			/clone="IMAGE:4995065"
			/cell_line="MG36"
			/lab_host="DH10B"
			/clone_1fb="Nih MGC 10"
			/note="Organ: cervix; Vector: pCMV-SPORT6; Site 1: NotI;
			Site 2: SalI; Cloned unidirectionally. Primer: oligo dT
			Average insert size 1.5 kb. Library prepared by Life
			Technologies."
BASE COUNT	238 a	196 c	286 g
ORIGIN			132 t
Alignment Scores:			
Pred. No.:	5..2e-57	Length:	852
Score:	1120..00	Matches:	226
Percent Similarity:	98.70%	Conservative:	1
Best local similarity:	98..264	Mismatches:	3
Query Match:	58..644	Indels:	3
Df:	12	Gaps:	0
US-09-972-758A-2 (1-359) x B1091005 (1-852)			
Oy	130	LouG1AlAProAlalag1yglugluglntgpglynglnglnaraglntleu	149
Db	5	GTGGGGCTTCCTCCCGCAGGGGCCAAAGAAGTGGACACGACGACGACTGGG	64
Oy	150	LYELyelyshisargtgrprproserlylsylsYArghistPLysPrcTYrTYLys	168
Db	65	AMGAAAAATATAGAGAGCGCCGTCCAAGAAAGCGGCA-TGGAAACCTACTACAAG	122
Oy	170	LeuThTrpGlunluYelyshelapgluyvsgInsertleuArgAlaserArgile	189
Db	124	CTGACCTGGGAAGAAAGAAAGTTTGACGAGAAACAGCCTTCGAGTTCAAAGATC	183
Oy	190	ArgAlaglumePheAlalySglygnPrOVA1AlAProyAsenThrThginPheUeu	209
Db	184	CGAGCCGAGATGTTCGCCCAAGGCCAGCCGGTCGGCCCTTAACACACGCAATTCTC	243
Oy	210	MetAspAsphisaSpInglunlpuprobleuYsthrGlyleUTyrSerlyArgAla	229
Db	244	ATGATGATCATCACGACGAGAGCGCGATCTCAAAAACGGGCTGTACTCCAACGGGCC	303
Oy	230	AlAlAlalyserAspAsphrserAspAspPhmetGluugluGlyglunluasp	249
Db	304	GCCGCAAAATCCACACACACACGACATGACGCTTCATGAAAGAGGGGATGAGAGAT	363
Oy	250	GlyglYserAspplYmetGlyglYAspGlySerGluPheLenglnArspPheSerGlu	263
Db	364	GGGGGACGATGGATGGAGGGGACGGCAGCGACTTCGAG	422
Oy	270	ThrTyrgluArgTyristhrgIuserieuglnaemetSerlyegInguJleuJleys	289
Db	423	ACGACGAGGGATACACACGAGAGCCTGCAGAACATGACGACGAGGACCTATCAAG	483
Oy	290	GluTYrlengluJleugluYscYseUsesraYmetGluarpGluAnaAsnArgleuArg	303
Db	483	GAGTACCTGGAACTGGAGAGATGCTCTCGCCATGAGAGACGAGAACMACCGGCTCGG	542
Oy	310	LeuGlyserlysrArgleuGlyglYAspAspa1AratVA1ArsgJuleuGluJleu	329
Db	543	CTGAGAGCAGCGGCTGGGTGGCCACACGCGCTGTCCGGGAGCTGGAGCTGGAGCTG	600
Oy	330	AspArgleuArgAlaJleuJleuJleuJleuJleuJleuJleuJleuJleuJleuJleu	349
Db	603	GACCGGCTGGCGCCGAGAACTCCACGCTGTGACCGAGAAACGAATCTGACG-CACAG	661
Oy	350	GluArgAlaProleuserlysrPheglYasp	359
Db	662	GAGGAGCGCGCTTTCAGATTGGAGC	691

RESULT	12
LOCUS	AL578242.C
DEFINITION	AL578242 Homo sapiens HELA CELLS COT 25-NORMALIZED Homo sapiens mRNA bp linear EST 01-JUN-2001
ACCESSION	CDS0DK002YF07 3-PRIME, mRNA sequence.
VERSION	AL578242
KEYWORDS	AL578242.2 GI:31316460
SOURCE	EST.
ORGANISM	Homo sapiens (human)
REFERENCE	Homo sapiens
AUTHORS	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
TITLE	Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
JOURNAL	Li W.B., Gruber,C., Jessee,J. and Polayes,D. Full-length cDNA libraries and normalization Unpublished
COMMENT	On Feb 16, 2001 this sequence version replaced gi:12942132. Contact: Genoscope Genoscope - Centre National de Sequencage BP 191 91006 EVRY cedex - France Email: seqref@genoscope.cns.fr Web : www.genoscope.cns.fr Library was constructed by Life Technologies, a division of Invitrogen. This sequence belongs to sequence cluster 8422.r For more information about this cluster, see <a href="http://www.genoscope.cns.fr/cgi-bin/cluster.cgi?seq=CSDK002C04NPIC&amp;cluster=8422.r">http://www.genoscope.cns.fr/ cgi-bin/cluster.cgi?seq=CSDK002C04NPIC&amp;cluster=8422.r</a> . Contact : Peng Liang Email : <a href="mailto:fliang@life-tech.com">fliang@life-tech.com</a> URL : <a href="http://fulllength.invitrogen.com/">http://fulllength.invitrogen.com/</a> Invitrogen Corporation 1600 Faraday Avenue Genoscope sequence ID : CSDK002CC04NPI. Location/Qualifiers 1..1078 /organism="Homo sapiens" /mol_type="mRNA" /db_xref="taxon:9606" /clone="CSODK002YF07" /cell_type="HELA CELLS COT 25-NORMALIZED" /cell_line="HELA" /clone_id="Homo sapiens HELA CELLS COT 25-NORMALIZED" /note="1st strand cDNA was primed with a NotI-oligo(dT) primer. Five prime end enriched, double-strand cDNA was digested with Not I and cloned into the Not I and EcoR V sites of the pCMVSPORT 6 vector. Library was normalized." BASE COUNT 228 a 310 c 232 g 290 t 18 others ORIGIN Alignment Scores: Pred. NO.: 6.31e-57 Length: 1078 Score: 1120.00 Matches: 219 Percent Similarity: 95.63% Conservative: 0 Best Local Similarity: 95.63% Mismatch: 10 Query Match: 58.64% Indels: 1 DB: Gaps: 0 US-09-972-758A-2 (1-359) x AL578242 (1-1078) Qy 131 GLVALPAPROIALAGLYGVLGLUGLUTPTPGYLINGNLGNARYGLINLEUPLYYS 150 Db 1060 GGCGCTCTTGCCSCAGGGGSGCAAGAAGAGGTGGGAACAAGCWYAGACAGCTGGGAAAG 1001+ Qy 151 LysLVSHIATGARARProSerLySVISVAARGHSITPDLYSPROTyrTYrLysLeu 170 Db 1000 AAAAMACATTAGGAGCGGCCGTCACNAAGAAAGCGGCATTGGAAACCCTACTACAAAGCTG 941 Qy 171 ThRTTPGIUGLUtLVSLySlySPheASpGIULuySGInserLeuArgAlaserAgIIleaRG 190 Db 940 ACCTGGGAAGAGAAAATAAGTTGCACGAGAAACAGACCTTTCGAGCTTCAAAGATCCGA 881 Qy 191 ALaGUmeCPheaAlalyGYlNglnProValAlapToYrAsnThTrThGlPhHeuWec 210 Db 880 GCCCGAGATGTTGCCCAAGGGCACGCGGTGGCCCTTAACACACACGACGATTCCTCATG 821 Qy 211 AspaSpHISAapGINGLUGluProaspLeuSThrGlyLeuTyrseryLysArgAlala 230 

[illegible]

## Alignment Scores:

Pred. No.: 1,68e-56 Length: 871  
 Score: 1111.50 Matches: 240  
 Percent Similarity: 90.07% Conservative: 5  
 Best Local Similarity: 88.24% Mismatches: 15  
 Query Match: 58.19% Indels: 15  
 DB: 10 Gaps: 2

US-09-972-758a-2 (1-359) x BG291150 (1-871)

```

QY 60 GYPRGGLUGLYGLUGLYSERLEUGLUSERGLNPROPROLEUGLNTHRGINALCYG 79
DB 11 GGGGCGGAGGGGAGAGGAGCGCTGGAGATCCCAACCCCTTGCAGACCCAGGCTGT 70
QY 80 PROGLUSERSERCYALEUARGUGLYGLUGLYGLVGLVGLVGLVGLVGLVGLVGLVGLV 99
DB 71 CCAGAACTTCACTGCTGCTGAGAGAGGCGGAGAGGCGGAGAGGAGAGAGAGAGAG 130
QY 100 GYGLYASPPHSPHSPHSPHSPHSPHSPHSPHSPHSPHSPHSPHSPHSPHSPHSPH 119
DB 131 GGGGCGGAGCTTCCCGCGCGCGGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 190
QY 120 GINPROCYSHISASPSERGLUALASERLYSLEUGLYALAPROALAGLYGLUGLU 139
DB 191 CAGCCTTGTCACTGACTCCAGAGCCAGTAAGTG-GGGGCTCTGCGCCGAGGGGCGAAG 249
QY 140 GUTTPGLVGLVGLVGLVGLVGLVGLVGLVGLVGLVGLVGLVGLVGLVGLVGLV 159
DB 250 GAGTGGGAGCAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 309
QY 160 LVSLYSLARHISTRYLSPRYLYLYLSLEUTHTRPGLUGLYSLYSLYSLYSLYSL 179
DB 310 MAGAAGCGGCA-TGMAAACCTTCTCAAGCTCACTGGAGAGAGAGAGAGAGAGAG 368
QY 180 GLUYSGINSELEUARGALASERARGILEARGALGLUWETHEALALYSGLYINPRO 199
DB 369 GAGAAACAGAGCCTTCAGACTTCAGAGATCCGAGCGAGATGTTCCCAAGGCGCAG 428
QY 200 VALALAIPROTYRASNTHRTHRGINPHELUWETASPSHISASPSGLNGLUGLUPROA 219
DB 429 GTCGCGCCCTATTAACAACAAGCGATTCCTCATGATGATCAACAAGAGAGAGAG 488
QY 220 LEULYTHGLYLEUTYRSELYSARGALALALALYSESERASPSAPHTHSEASPA 239
DB 489 CTCMAAACCGGCTGTACTCAAGCGGCGC--GCCAATCCACCAACACACGAGTAC 546
QY 240 ASPHMETGLUGLYGLUGLYGLUGLYGLUGLYGLUGLYGLUGLYGLUGLYGLUGLY 259
DB 547 GACTTCATGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 606
QY 259 YSERGLUPHELEUGLARGHSPHSESEGLUTHTYRGLU-ARGLYRHTHTGLUSE 279
DB 607 CAGCGAGTTCGACACGAGGAGCTTCTCGAGAGCTGAGCCGCTGACCAACGAGAG 666
QY 279 EUGLIN-ASPMETSERLYSGIN-GLUENILEYLS-GLUTYRLEUGLUU-GLUYLS 296
DB 667 TGCAGAGAGTGAAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 726
QY 297 CYLSLEUSERARGMETGLUASPMASPMASPMASPMASPMASPMASPMASPMAS 311
DB 727 GCTCTCGGGGCGCTGAGAGGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 786
QY 312 SERLYSARGLEUGLYGLYASPA 319
DB 787 AGACCCCGTTTGGAGCGGGGAG 810

```

RESULT 14  
 BM474968 1294 bp mRNA linear EST 05-FEB-2002  
 LOCUS AGENCOUNT\_6476735 NIH\_MGC\_88 Homo sapiens cDNA IMAGE:5562828  
 DEFINITION 5', mRNA sequence.  
 ACCESSION BM474968

VERSION BM474968.1 GI:18524010  
 KEYWORDS EST.  
 SOURCE Homo sapiens (human)  
 ORGANISM Homo sapiens

REFERENCE  
 AUTHORS NIH-MGC  
 TITLE NIH-MGC  
 JOURNAL NIH-MGC  
 COMMENT NIH-MGC  
 Email: cgapbs-remail.nih.gov  
 Tissue Procurement: ATCC  
 CDNA Library Preparation: Life Technologies, Inc.  
 DNA Sequencing by: The I.M.A.G.E. Consortium (LNL)  
 Clone distribution: MGC clone distribution information can be found through the I.M.A.G.E. Consortium/LNL at:  
 http://image.llnl.gov  
 Plate: L1292  
 Location/Qualifiers  
 1. 1294

FEATURES  
 source  
 1. 1294  
 /organism="Homo sapiens"  
 /mol\_type="mRNA"  
 /db\_xref="taxon:9606"  
 /clone="IMAGE:5562828"  
 /tissue\_type="duodenal adenocarcinoma, cell line"  
 /lab\_host="DH10B (phage-resistant)"  
 /clone\_lib="NIH\_MGC\_88"  
 /note="Organ: small intestine; Vector: pCMV-SPORT6; Site 1: NotI; Site 2: SalI; Cloned unidirectionally; oligo-dT primed. Average insert size 1.767 kb. Library enriched for full-length clones and constructed by Life Technologies. Note: this is a NIH MGC Library."

BASE COUNT 263 a 380 c 407 g 241 t 3 others

## Alignment Scores:

Pred. No.: 3.79e-54 Length: 1294  
 Score: 1074.00 Matches: 214  
 Percent Similarity: 89.80% Conservative: 6  
 Best Local Similarity: 87.35% Mismatches: 14  
 Query Match: 56.23% Indels: 11  
 DB: 12 Gaps: 3

US-09-972-758a-2 (1-359) x BM474968 (1-1294)

```

QY 31 GUATGPROGLYALAGLUGLUARGYALPROGLUGLUASPSERARGTRPGLINSE 50
DB 3 GAGCGCGCCCGAGAGGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 62
QY 51 ALAPHEPROGLNLEUGLYGLYARGPROGLYPROGLUGLYGLUGLYSERLEUGL 70
DB 63 GCGTTCGCCCACTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 122
QY 71 PROPROPROLEUGLTHRGALACYSPPROGLUSERSERCYALEUARGUGLYGLUY 90
DB 123 CCACTCCCTTCAGACCCAGAGCTGTCCAGAACTTAGCTGAGAGAGAGAGAGAG 182
QY 91 GYGLNAGNGLYASPSERSEALAGLYGLYASPPHSPHSPHSPHSPHSPHSPHSP 110
DB 183 GGCAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 242
QY 111 PROTHRPROGLUALAGLULEULEUALAGLNPQCYSHISASPSERGLUALASER 130
DB 243 CCGAGCGCCGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 302
QY 131 GYALAPROALALAGLYGLUGLUGLUGLUTPRGLVGLNGLVGLVGLVGLVGLV 150
DB 303 GGGGCTCTCGCGCAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 362
QY 151 LVSLYSLARHISTRYLSPRYLYLYLSLEUTHTRPGLUGLYSLYSLYSLYSLY 170

```







GenCore version 5.1.6  
Copyright (c) 1993 - 2004 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: February 5, 2004, 12:03:48 ; Search time 81 Seconds

(without alignments)  
703.492 Million cell updates/sec

Title: US-09-972-758a-2

Perfect score: 1910  
Sequence: 1 MAEPFLSEYQHQPQNSNCTG.....LTENELHROERAPLRFKFD 359

Scoring table: BLOSUM62  
Gapop 10.0 , Gapext 0.5

Searched: 1107863 seqs, 158726573 residues.

Total number of hits satisfying chosen parameters: 1107863

Minimum DB seq length: 0  
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%  
Maximum Match 100%

Listing first 45 summaries

Database :  
A.Geneseq.19jun03.\*  
1: /SIDSI/gcgdata/geneseq/geneseq-emb1/AA1980.DAT.\*  
2: /SIDSI/gcgdata/geneseq/geneseq-emb1/AA1981.DAT.\*  
3: /SIDSI/gcgdata/geneseq/geneseq-emb1/AA1982.DAT.\*  
4: /SIDSI/gcgdata/geneseq/geneseq-emb1/AA1983.DAT.\*  
5: /SIDSI/gcgdata/geneseq/geneseq-emb1/AA1984.DAT.\*  
6: /SIDSI/gcgdata/geneseq/geneseq-emb1/AA1985.DAT.\*  
7: /SIDSI/gcgdata/geneseq/geneseq-emb1/AA1986.DAT.\*  
8: /SIDSI/gcgdata/geneseq/geneseq-emb1/AA1987.DAT.\*  
9: /SIDSI/gcgdata/geneseq/geneseq-emb1/AA1988.DAT.\*  
10: /SIDSI/gcgdata/geneseq/geneseq-emb1/AA1989.DAT.\*  
11: /SIDSI/gcgdata/geneseq/geneseq-emb1/AA1990.DAT.\*  
12: /SIDSI/gcgdata/geneseq/geneseq-emb1/AA1991.DAT.\*  
13: /SIDSI/gcgdata/geneseq/geneseq-emb1/AA1992.DAT.\*  
14: /SIDSI/gcgdata/geneseq/geneseq-emb1/AA1993.DAT.\*  
15: /SIDSI/gcgdata/geneseq/geneseq-emb1/AA1994.DAT.\*  
16: /SIDSI/gcgdata/geneseq/geneseq-emb1/AA1995.DAT.\*  
17: /SIDSI/gcgdata/geneseq/geneseq-emb1/AA1996.DAT.\*  
18: /SIDSI/gcgdata/geneseq/geneseq-emb1/AA1997.DAT.\*  
19: /SIDSI/gcgdata/geneseq/geneseq-emb1/AA1998.DAT.\*  
20: /SIDSI/gcgdata/geneseq/geneseq-emb1/AA1999.DAT.\*  
21: /SIDSI/gcgdata/geneseq/geneseq-emb1/AA2000.DAT.\*  
22: /SIDSI/gcgdata/geneseq/geneseq-emb1/AA2001.DAT.\*  
23: /SIDSI/gcgdata/geneseq/geneseq-emb1/AA2002.DAT.\*  
24: /SIDSI/gcgdata/geneseq/geneseq-emb1/AA2003.DAT.\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

#### SUMMARIES

Result No.	Score	Query Match	Length DB	ID	Description
1	1910	100.0	359	19	AAW85455
2	1910	100.0	359	23	ABP61799
3	1910	100.0	359	23	ABP76495
4	1903	99.6	359	22	ABP95465
5	545.5	28.6	23	22	ABP43855
6	308.5	16.2	152	22	AAW25726
7	237	12.4	134	22	ABG15306
8	209	10.9	349	22	ABBS9958
9	155.5	8.1	281	22	ABG15303
					Secreted protein e
					Human polypeptide
					Human oestrogen do
					Human protein segu
					Mouse putative pro
					Human protein segu
					Novel human diagno
					Drosophila melanog
					Novel human diagno

10	147	7.7	954	22	AAU14615
11	143	7.5	288	22	ABG20362
12	141.5	7.4	1192	22	ABBS9642
13	138	7.2	557	19	AAU20666
14	135.5	7.1	530	21	AAU94461
15	135.5	7.1	530	21	AAU74503
16	135.5	7.1	1749	22	ABG00839
17	135	7.1	554	23	ABG04721
18	135	7.1	555	22	AAU93869
19	134.5	7.0	510	22	ABH11764
20	134.5	7.0	510	22	AAW79741
21	134.5	7.0	546	22	AAW82808
22	134.5	7.0	639	24	ABR41186
23	134.5	7.0	639	21	ABR41210
24	133	7.0	538	21	AAU10043
25	133	7.0	648	20	AAU17946
26	133	7.0	1737	21	AAU10044
27	132.5	6.9	374	22	AAW64026
28	132.5	6.9	530	20	AAW97775
29	132.5	6.9	530	23	AAU74501
30	132.5	6.9	530	22	AAU74502
31	132	6.9	528	22	AAW78918
32	131	6.9	932	22	ABBS2556
33	131	6.9	1424	22	AAW39253
34	131	6.9	1464	22	AAW41039
35	131	6.9	1898	20	AAU30795
36	130.5	6.8	819	22	AAW94316
37	130	6.8	611	20	AAU29039
38	130	6.8	611	22	AAU25510
39	129.5	6.8	1743	22	ABG10928
40	128	6.7	670	23	ABP62902
41	127.5	6.7	984	22	AAO12986
42	127	6.6	654	22	ABBS3266
43	126.5	6.6	470	24	AAE33670
44	126.5	6.6	661	22	ABBS1881
45	126.5	6.6	710	22	ABG20363

#### ALIGNMENTS

RESULT 1	AAW85455	AAW85455 standard; Protein; 359 AA.
ID	AAW85455	
XX	AAW85455	
AC	AAW85455	
XX		
DT	25-FEB-1999	(first entry)
XX		
DE	Secreted protein encoded by clone bp783_3.	
XX		
KW	Secreted protein; nutritional activity; immune stimulating; vaccine;	
KW	suppressing activity; haematopoiesis regulating activity;	
KW	tissue growth activity; activin; inhibin activity; chemotaxis;	
KW	chemokinetic activity; haemostasis; thrombolytic activity; receptor;	
KW	ligand; anti-inflammatory; cadherin; tumour invasion suppressor;	
XX	tumour inhibition; gene therapy.	
OS	Homo sapiens.	
XX		
PN	WO9842739-A2.	
XX		
PD	01-OCT-1998.	
XX		
EF	20-MAR-1998;	98WO-US05653.
XX		
PR	19-MAR-1998;	98US-0044466.
XX		
PR	21-MAR-1997;	97US-0822167.
XX		
PA	(GENV ) GENETICS INST INC.	
XX		
PI	Agostino M, Jacobs K, Lavallie ER, McCoy JM, Merberg D;	
XX	Racie LA, Spaulding V, Treacy M;	

XX MPI; 1998-609890/51.  
 DR N-PSDB; AAV82778.  
 XX  
 PT New polynucleotides encoding secreted human proteins - derived from  
 PT human foetal brain, adult brain, foetal kidney, placenta or adult  
 PT pineal gland cDNA libraries.  
 PS  
 XX  
 PS Claim 1; Page 67-68; 113pp; English.  
 XX  
 CC The present sequence represents a secreted protein. The polynucleotide  
 CC and secreted protein are predicted to have biological activities which  
 CC would make them suitable for treating, preventing or ameliorating medical  
 CC conditions in humans and animals, although no supporting data is given.  
 CC Suggested activities include nutritional activity, immune stimulating  
 CC (e.g. as vaccines) or suppressing activity, haematopoiesis regulating  
 CC activity, tissue growth activity, activin/inhibin activity,  
 CC chemotactic/chemokinetic activity, haemostatic and thrombolytic activity,  
 CC receptor/ligand activity, anti-inflammatory activity, cadherin/tumour  
 CC invasion suppressor activity, and tumour inhibition activity (no data is  
 CC given in the specification to support these activities). The  
 CC polynucleotide is also stated to be useful for gene therapy.  
 CC  
 XX  
 SQ Sequence 359 AA;  
 Query Match 100.0%; Score 1910; DB 19; Length 359;  
 Best Local Similarity 100.0%; Pred. No. 6.5e-153;  
 Matches 359; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
 QY 1 MAEPFLSEYOHQOTNSCTGAAGVQBELNPERPPGAEERVPEDSRWOSRAFPQLGGRPG 60  
 DB 1 MAEPFLSEYOHQOTNSCTGAAGVQBELNPERPPGAEERVPEDSRWOSRAFPQLGGRPG 60  
 QY 61 PEEGSGLESOPPLQTOQAPRESSCLREGEKQNGDSSAGDPPPAEVPTEAEALLQ 120  
 DB 61 PEEGSGLESOPPLQTOQAPRESSCLREGEKQNGDSSAGDPPPAEVPTEAEALLQ 120  
 QY 121 PCGDSASKLGAAPAGAEEMGQOQOLGKKRRRRPSSKKRWMKPYKLTWEKKKFPDE 180  
 DB 121 PCGDSASKLGAAPAGAEEMGQOQOLGKKRRRRPSSKKRWMKPYKLTWEKKKFPDE 180  
 QY 181 KQSLRASRIAPAEPAKQPPAPYNTTQFLMDHDQEBPDLKTLYSKRAAKSDDTSDDD 240  
 DB 181 KQSLRASRIAPAEPAKQPPAPYNTTQFLMDHDQEBPDLKTLYSKRAAKSDDTSDDD 240  
 QY 241 FMEEGSEBPGGSGMGDGGSEFLQRPSEFYERYHTESLQMSKQELIKYLEKCLSR 300  
 DB 241 FMEEGSEBPGGSGMGDGGSEFLQRPSEFYERYHTESLQMSKQELIKYLEKCLSR 300  
 QY 301 MEDENNRRLRESKRLGDDARVALELELDRLRAENLQULTENELHQOERAPLSKFGD 359  
 DB 301 MEDENNRRLRESKRLGDDARVALELELDRLRAENLQULTENELHQOERAPLSKFGD 359  
 RESULT 2  
 ABP61799  
 ID ABP61799 standard; Protein; 359 AA.  
 XX  
 AC ABP61799;  
 XX  
 DT 04-OCT-2002 (first entry)  
 DE Human polypeptide SEQ ID NO 153.  
 XX  
 XX Human; cytostatic; antirheumatic; antiarthritic; vulnery; analgesic;  
 KW antiinflammatory; antibacterial; immunosuppressive; antiparkinsonian;  
 KW neuroprotective; nootropic; osteopathic; haemostatic; vasotropic;  
 KW anticancer; fungicide; antidiabetic; antiaesthetic; antiallergic;  
 KW immunostimulant; antiparasitic; secreted protein; transmembrane protein;  
 KW cytokine; cell proliferation; cell differentiation; autoimmune disease;  
 KW stem cell; growth factor; nervous system disease; neuropathy;  
 KW Alzheimer's disease; Parkinson's disease; Huntington's disease;  
 KW osteoporosis; severe combined immunodeficiency; SCID; infection;

KW multiple sclerosis; rheumatoid arthritis; gene therapy.  
 XX  
 OS Homo sapiens.  
 XX  
 PN US2002065394-A1.  
 XX  
 PD 30-MAY-2002.  
 XX  
 PF 22-DEC-2000; 2000US-0745763.  
 XX  
 PR 18-MAR-1998; 98US-0040963.  
 XX  
 PA (JACO/) JACOBS K.  
 PA (MCCO/) MCCOY J M.  
 PA (LAVA/) LAVALLIE E R.  
 PA (COLL/) COLLINS-RACIE L A.  
 PA (EVAN/) EVANS C.  
 PA (MERB/) MERBERG D.  
 PA (TREX/) TREACY M.  
 PA (SPA/) SPAULDING V.  
 XX  
 PI Jacobs K, McCoy JM, LaVallie ER, Collins-Racie LA, Evans C;  
 PI Merberg D, Treacy M, Spaulding V;  
 DR MPI; 2002-582343/62.  
 DR N-PSDB; ABQ92015.  
 XX  
 PT Novel secreted or transmembrane protein and polynucleotide encoding the  
 PT protein, useful for diagnosis and treatment of neurological disorders,  
 PT cancer, autoimmune diseases, bone disorders and lung or liver fibrosis  
 PT  
 PS Claim 44; Page 112-113; 284pp; English.  
 XX  
 CC The invention relates to human secreted or transmembrane protein (I),  
 CC their fragments and is encoded by specific complementary deoxyribonucleic  
 CC acid (cDNA) inserts (II), where the protein is substantially free from  
 CC other mammalian proteins. (I) are useful for preventing, treating or  
 CC ameliorating a medical condition, especially immunological treatment or  
 CC prevention of tumours. (I) exhibits activity relating to angiogenesis,  
 CC cytokine, cell proliferation, cell differentiation, antinflammatory,  
 CC stem cell growth factor activity and activin or inhibin-related  
 CC activities. (I) can be used to manipulate stem cells in culture to give  
 CC rise to neuroepithelial cells that can be used to augment or replace  
 CC cells damaged by illness, autoimmune disease, accidental damage or  
 CC genetic disorders. (I) induces the proliferation of neural cells and  
 CC regeneration of nerve and brain tissue and is useful for the treatment of  
 CC central and peripheral nervous system diseases and neuropathies, such as  
 CC Alzheimer's, Parkinson's disease, Huntington's disease, amyotrophic  
 CC lateral sclerosis. (I) is involved in chemotactic or chemokinetic  
 CC activity, regulation of haematopoiesis and is useful for treating myeloid  
 CC or lymphoid cell disorders, platelet disorders such as thrombocytopenia  
 CC and for regeneration of bone, cartilage, tendon, ligament and/or nerve  
 CC tissue growth and in tissue repair, healing of burns, incisions, ulcers,  
 CC for treating osteoporosis, osteoarthritis, bone degenerative disorders or  
 CC periodontal disease. (II) is also useful for gut protection or  
 CC regeneration and treatment of lung or liver fibrosis, reperfusion injury  
 CC in various tissues, various immune deficiencies and disorders including  
 CC severe combined immunodeficiency (SCID), bacterial or fungal infections,  
 CC autoimmune disorders e.g. multiple sclerosis, rheumatoid arthritis,  
 CC diabetes mellitus, myasthenia gravis, allergic reactions and conditions,  
 CC such as asthma or other respiratory problems. (II) is useful to express  
 CC recombinant protein, as markers for tissues in which the corresponding  
 CC protein is preferentially expressed and in gene therapy. The present  
 CC sequence is that of a polypeptide of the invention.  
 XX  
 SQ Sequence 359 AA;  
 Query Match 100.0%; Score 1910; DB 23; Length 359;  
 Best Local Similarity 100.0%; Pred. No. 6.5e-153;  
 Matches 359; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
 QY 1 MAEPFLSEYOHQOTNSCTGAAGVQBELNPERPPGAEERVPEDSRWOSRAFPQLGGRPG 60

Db 1 MABPFLSEYQHOPQTSNCTGAAAVOELNPERPPGAEEVPEDSRWQSRAPFOLGGRPG 60  
QY 61 PEGEGSLSEOPPLQTOACPESSCLREGEKQNGDSSAGDPPPAEVEPTPEALLAQ 120  
Db 61 PEGEGSLSEOPPLQTOACPESSCLREGEKQNGDSSAGDPPPAEVEPTPEALLAQ 120  
QY 121 PCHDSASAKLGAPAAAGGEEWQOQOLGKKRRRRPSSKKHMKRYYLKLTWEKKKFPDE 180  
Db 121 PCHDSASAKLGAPAAAGGEEWQOQOLGKKRRRRPSSKKHMKRYYLKLTWEKKKFPDE 180  
QY 181 KOSLRASIRIAEMFAGQPVAPYNTTQFLMDHDHDEPDLKTLGYSKRAAKSDDTSDDD 240  
Db 181 KOSLRASIRIAEMFAGQPVAPYNTTQFLMDHDHDEPDLKTLGYSKRAAKSDDTSDDD 240  
QY 241 FMEEGEEDGSGDMGDSSEFLQDFSETYRYHTESLQNNMSKQELIKEYLEKCLSR 300  
Db 241 FMEEGEEDGSGDMGDSSEFLQDFSETYRYHTESLQNNMSKQELIKEYLEKCLSR 300  
QY 301 MEDENNRRLRESKRLGGDDARVRELELELDRLRAENLQLTENELHROQERAPLSKFGD 359  
Db 301 MEDENNRRLRESKRLGGDDARVRELELELDRLRAENLQLTENELHROQERAPLSKFGD 359

RESULT 3  
ABB76495  
ID ABB76495 standard; Protein; 359 AA.  
AC ABB76495;

XX 10-SEP-2002 (first entry)  
XX Human oestrogen downregulated gene EDG1 protein.  
XX EDG1; oestrogen downregulated gene; tumour suppressor; human;  
XX breast cancer; prostate cancer; testicular cancer; ovarian cancer;  
XX uterine cancer; colon cancer; gene therapy.  
XX Homo sapiens.  
XX Key Location/Qualifiers  
XX Region 150..177  
XX /note= "nuclear localisation signal"  
XX Misc-difference 154  
XX /note= "encoded by AAG"  
XX Misc-difference 171  
XX /note= "encoded by AAC"  
XX W0200228879-A1.  
XX 11-APR-2002.  
XX 05-OCT-2001; 2001WO-US31300.  
XX 05-OCT-2000; 2000US-238187P.  
XX (UYCA-) UNIV CASE WESTERN RESERVE.  
XX Montano M, Wittman B;  
XX WPI; 2002-519107/55.  
XX N-PSDB; ABN84013.  
XX Polynucleotides encoding Estrogen Down-Regulated Gene 1 proteins,  
XX useful for the prevention, diagnosis and treatment of e.g. breast  
XX cancer, testicular cancer, prostate cancer, uterine cancer, cervical  
XX cancer, ovarian cancer and colon cancer -  
XX Claim 7, Fig 1A-B; 52pp; English.  
XX The present sequence is the protein sequence of human EDG1, a  
XX 40 kDa protein encoded by the tumour suppressor gene designated  
XX human oestrogen downregulated gene 1 (see ABN84013). EDG1 mRNA

CC expression is prevalent in normal mammary epithelial cells and in  
CC other human hormone-responsive tissues such as the ovary, prostate  
CC and testis. Expression is low in breast cancer epithelial cells.  
CC Oestradiol, which induces breast cancer cell growth, has an  
CC inhibitory effect on EDG1 mRNA expression in breast cancer cells.  
CC Hexamethylene bis-acetamide, an inducer of differentiation and  
CC apoptosis, upregulates EDG1 mRNA expression in breast cancer cells.  
CC The invention provides EDG1 polynucleotides and polypeptides. In  
CC a claimed method of detecting cancerous cells, a test sample from  
CC an individual suspected of having, or known to have breast,  
CC testicular, prostate, uterine, cervical, ovarian or colon cancer is  
CC contacted with an anti-EDG1 antibody. A decrease in the level  
CC or antigen-antibody complex compared to the level of a control  
CC sample indicates cancerous cells. A claimed method for decreasing  
CC the proliferation of breast, prostate, testicular, ovarian, uterine,  
CC cervical or colon cancer cells involves increasing EDG1 protein  
CC activity in the cells, either by contacting the cells with EDG1  
CC protein, its fragment or functional equivalent, or with a nucleic  
CC acid encoding EDG1 protein, its fragment or functional equivalent.  
XX Sequence 359 AA;

Query Match 100.0%; Score 1910; DB 23; Length 359;  
Best Local Similarity 100.0%; Pred. No. 6.5e-153;  
Matches 359; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 MABPFLSEYQHOPQTSNCTGAAAVOELNPERPPGAEEVPEDSRWQSRAPFOLGGRPG 60  
Db 1 MABPFLSEYQHOPQTSNCTGAAAVOELNPERPPGAEEVPEDSRWQSRAPFOLGGRPG 60  
QY 61 PEGEGSLSEOPPLQTOACPESSCLREGEKQNGDSSAGDPPPAEVEPTPEALLAQ 120  
Db 61 PEGEGSLSEOPPLQTOACPESSCLREGEKQNGDSSAGDPPPAEVEPTPEALLAQ 120  
QY 121 PCHDSASAKLGAPAAAGGEEWQOQOLGKKRRRRPSSKKHMKRYYLKLTWEKKKFPDE 180  
Db 121 PCHDSASAKLGAPAAAGGEEWQOQOLGKKRRRRPSSKKHMKRYYLKLTWEKKKFPDE 180  
QY 181 KOSLRASIRIAEMFAGQPVAPYNTTQFLMDHDHDEPDLKTLGYSKRAAKSDDTSDDD 240  
Db 181 KOSLRASIRIAEMFAGQPVAPYNTTQFLMDHDHDEPDLKTLGYSKRAAKSDDTSDDD 240  
QY 241 FMEEGEEDGSGDMGDSSEFLQDFSETYRYHTESLQNNMSKQELIKEYLEKCLSR 300  
Db 241 FMEEGEEDGSGDMGDSSEFLQDFSETYRYHTESLQNNMSKQELIKEYLEKCLSR 300  
QY 301 MEDENNRRLRESKRLGGDDARVRELELELDRLRAENLQLTENELHROQERAPLSKFGD 359  
Db 301 MEDENNRRLRESKRLGGDDARVRELELELDRLRAENLQLTENELHROQERAPLSKFGD 359

RESULT 4  
AAB95465  
ID AAB95465 standard; Protein; 359 AA.  
AC AAB95465;  
XX 26-JUN-2001 (first entry)  
XX Human protein sequence SEQ ID NO:17953.  
XX Human, primer; detection; diagnosis; antisense therapy; gene therapy.  
XX Homo sapiens.  
XX BP1074617-A2.  
XX 07-FEB-2001.  
XX 28-JUL-2000; 2000EP-0116126.  
XX 29-JUL-1999; 99JP-0248036.  
XX 27-AUG-1999; 99JP-0300253.

PR 11-JAN-2000; 2000JP-0118776.  
PR 02-MAY-2000; 2000JP-0183767.  
PR 09-JUN-2000; 2000JP-0241899.

XX (HELI-) HELIX RES INST.

PI Oca T, Isogai T, Nishikawa T, Hayashi K, Saito K, Yamamoto J;  
PI Ishii S, Sugiyama T, Wakamatsu A, Nagai K, Otsuki T;

XX MPI; 2001-318749/34.

PT Primer sets for synthesizing polynucleotides, particularly the 5602  
PT full-length CDNA defined in the specification, and for the detection  
PT and/or diagnosis of the abnormality of the proteins encoded by the  
PT full-length CDNA -

PS Claim 8; SEQ ID 17953; 2537pp + CD ROM; English.

CC The present invention describes primer sets for synthesizing 5602  
CC full-length CDNA defined in the specification. Where a primer set  
CC comprises: (a) an oligo-dT primer and an oligonucleotide complementary  
CC to the complementary strand of a polynucleotide which comprises one of  
CC the 5602 nucleotide sequences defined in the specification, where the  
CC oligonucleotide comprises at least 15 nucleotides; or (b) a combination  
CC of an oligonucleotide comprising a sequence complementary to the  
CC complementary strand of a polynucleotide which comprises a 5'-end  
CC sequence and an oligonucleotide comprising a sequence complementary to a  
CC polynucleotide which comprises a 3'-end sequence, where the  
CC oligonucleotide comprises at least 15 nucleotides and the combination of  
CC the 5'-end sequence/3'-end sequence is selected from those defined in  
CC the specification. The primer sets can be used in antisense therapy and  
CC in gene therapy. The primers are useful for synthesizing polynucleotides,  
CC particularly full-length CDNA. The primers are also useful for the  
CC detection and/or diagnosis of the abnormality of the proteins encoded by  
CC the full-length CDNA. The primers allow obtaining of the full-length  
CC CDNA easily without any specialised methods. AAH03166 to AAH3628 and  
CC AAH1633 to AAH18742 represent human cDNA sequences; AAH92446 to  
CC AAH95893 represent human amino acid sequences; and AAH13629 to AAH13632  
CC represent oligonucleotides, all of which are used in the exemplification  
CC of the present invention.

XX Sequence 359 AA;

Query Match 99.6%; Score 1903; DB 22; Length 359;

Best Local Similarity 99.7%; Pred. No. 2.6e-152;  
Matches 358; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 1 MABPFLSEYHOGQTSNCTGAAVQEBLNERRPPGAERPPEDSRKQSAFPOLGGRPG 60  
DB 1 MABPFLSEYHOGQTSNCTGAAVQEBLNERRPPGAERPPEDSRKQSAFPOLGGRPG 60  
QY 61 PEGGSLSEOPPLQOTACPESSCLREGEKQNGDDSSAGDPPPAVEPTBEALLAQ 120  
DB 61 PEGGSLSEOPPLQOTACPESSCLREGEKQNGDDSSAGDPPPAVEPTBEALLAQ 120  
QY 121 PCHDSEASKLGAAPAGGEEEMGOQOROLGKKRRRPPSKKKRHKPPYKLTWEKKKFFDE 180  
DB 121 PCHDSEASKLGAAPAGGEEEMGOQOROLGKKRRRPPSKKKRHKPPYKLTWEKKKFFDE 180  
QY 181 KOSLRASRIAEEMFAKQPVAPYNTTQFLMDHDHDEEPDLKTGLYSRAAASDDTSDDD 240  
DB 181 KOSLRASRIAEEMFAKQPVAPYNTTQFLMDHDHDEEPDLKTGLYSRAAASDDTSDDD 240  
QY 241 FMEGGEGEDGSGMGDGESEFLORDPSEYTERHYTESLONMSKOLIKLEYLEKCLSR 300  
DB 241 FMEGGEGEDGSGMGDGESEFLORDPSEYTERHYTESLONMSKOLIKLEYLEKCLSR 300  
QY 301 MEDENNRRLRESKRLGDDARVLELELDRLRAENLQLTENELHQERAPLSKFGD 359  
DB 301 MEDENNRRLRESKRLGDDARVLELELDRLRAENLQLTENELHQERAPLSKFGD 359

RESULT 5

ABP43855  
ID ABP43855 standard; Protein, 286 AA.

XX ABP43855;

XX 26-FEB-2003 (first entry)

DE Mouse putative protein #15.

XX Neuroprotective; immunomodulator; cancer;

KW cytostatic; anti-inflammatory; gene therapy; Huntington's disease;

KW wound, burn, ulcer; Alzheimer's disease; autoimmune disorder; inflammation;

KW vlnary.

XX Mus musculus.

XX WO00231111-A2.

XX 18-APR-2002.

XX 11-OCT-2001; 2001WO-US27760.

XX 12-OCT-2000; 2000US-0687527.

XX (HYSE-) HYSEQ INC.

PI Tang YT, Liu C, Zhou P, Asundi V, Zhang J, Zhao QA, Ren F;

PI Xue AJ, Yang Y, Weinman T, Drmanac RT;

XX N-PSDB; ABQ61099.

XX New polypeptides and their encoded proteins, useful as nutritional

PT sources or supplements, or in gene therapy, particularly for treating

PT wounds, Alzheimer's disease, amyotrophic lateral sclerosis, cancer or

PT inflammation -

XX Claim 20; SEQ ID # 758; 357pp + sequence listing; English.

XX The invention relates to 446 newly isolated polynucleotide sequences.

CC The activity of polynucleotides of the invention may be described as,

CC vlnary, neuroprotective, immunomodulator, cyostatic and

CC anti-inflammatory. Compositions comprising nucleic acids of the invention

CC are useful for treating a mammalian subject, or as nutritional sources or

CC supplements. These are useful in gene therapy, particularly for treating

CC wounds, burns or ulcers, Alzheimer's disease, Huntington's disease,

CC amyotrophic lateral sclerosis, autoimmune disorders, cancer or

CC inflammation. The nucleic acids and polypeptides are also useful in

CC diagnostic and research methods. The sequences given in records

CC ABP43544-ABP43989 represent polypeptides encoded by polynucleotides of

CC the invention.

CC NOTE: The sequence data for this patent did not form part of the printed

CC specification, but was obtained in electronic format directly from WIPO

CC at ftp.wipo.int/pub/published\_pct\_sequences.

XX Sequence 286 AA;

Query Match 28.6%; Score 545.5; DB 23; Length 286;

Best Local Similarity 43.8%; Pred. No. 8.4e-38;

Matches 128; Conservative 36; Mismatches 93; Indels 35; Gaps 6;

QY 73 PLQOTACPESSCLREGEKQNGDDSSAGDPPPAVEPTBEALLAQPC-- 122

DB 5 PNOTACNASSPYALBAKTSAGSPQTPRRHNDGGSPLPRHSHEDDDLAGAVG 64

QY 123 --HDSASKLGAAPAGGEEEMGOQOROLGKKRRRPPSKKKRHKPPYKLTWEKKKFFDE 180

DB 65 LGMNRSRPTQSPGCSAE-----AVLARKKRRRPPSKKKRHKPPYKLTWEKKKFFDE 118

QY 181 KOSLRASRIAEEMFAKQPVAPYNTTQFLMDHDHDEEP--DKTGLYSRAAASDDTSD 238

DB 119 ROSQRASRIAEEMFAKQPVAPYNTTQFLMDHDHDEEP--DKTGLYSRAAASDDTSD 173

QY 239 DFMEEGEGDGGSDGKMGDSEFLQRFSEYERYHTSLQNMKSQKELKEYLEKCL 298  
 DB 174 -----EAGSDGRGRAHGEFORDFSEYERFHTSLQGRSQKELVRDYLEKRL 224  
 QY 299 SRMEDENRRLRLSKRLGDDAR-VRELEELRLRAENQLTENEHMQ 349  
 DB 225 SQAEETRRRLQQLQACTGQSCROVELAEVORLRTENQRLROENQNMRE 276

RESULT 6  
 AAM25726  
 ID AAM25726 standard; Protein; 152 AA.  
 XX  
 AC AAM25726;  
 XX  
 DT 16-OCT-2001 (first entry)  
 XX  
 DE Human protein sequence SEQ ID NO:1241.  
 XX  
 KW Human; cancer; ulcer; HIV infection; human immunodeficiency virus;  
 KW anti-inflammatory; antirheumatic; antiarthritic; immunosuppressive;  
 KW antidiabetic; endocrine; cardiac; central nervous system; virucide;  
 KW anti-HIV; fungicide; antimutagen; cardiovascular; antianaemic; anaemia;  
 KW antileukemic; haemostatic; vulnery; antidiabetic; osteopathic; eczema;  
 KW dermatological; antiallergic; antiasthmatic; antiparkinsonian; infection;  
 KW neuroprotective; antidepressant; nootropic; antiparkinsonian; inflammation;  
 KW immunostimulant; gene therapy; antisense therapy; vaccine; inflammation;  
 KW antianaphylactic; rheumatoid arthritis; septic shock; pancreatitis;  
 KW cardiac dysfunction; neuropathology; cardiac anaphylaxis; autoimmunity;  
 KW genetic disease; haematopoietic disorder; platelet disorder; asthma;  
 KW thrombocytopaenia; osteoporosis; severe combined immunodeficiency;  
 KW allergic rhinitis; diabetes; multiple sclerosis; depression;  
 KW Alzheimer's disease; Parkinson's disease; neurodegenerative disorder;  
 KW neurological disorder.  
 XX  
 OS Homo sapiens.  
 XX  
 PN WO200153455-A2.  
 XX  
 PD 26-JUL-2001.  
 XX  
 PF 22-DEC-2000; 2000WO-US35017.  
 XX  
 PR 23-DEC-1999; 99US-0471275.  
 PR 21-JAN-2000; 2000US-0488725.  
 PR 25-APR-2000; 2000US-0552317.  
 XX  
 PA (HYSE-) HYSEQ INC.  
 XX  
 PI Tang YT, Liu C, Drmanac RT;  
 XX  
 DR WPI; 2001-457603/49.  
 DR N-PSDB; AAM99667.  
 XX  
 PT Isolated human polynucleotides encoding polypeptides, useful for the  
 PT treatment and diagnosis of e.g. cancer, ulcers and HIV infection -  
 XX  
 Claim 20; Page 257; 1217pp; English.

AAH99166 to AAH99904 encode the human proteins given in AAM25225 to  
 CC AAM25963. The proteins can have activities based on the tissues and  
 CC cells they are expressed in, such as: anti-inflammatory; antirheumatic;  
 CC antidiabetic; immunosuppressive; antidiabetic; endocrine; cardiac;  
 CC central nervous system; virucide; anti-HIV; fungicide; antimutagen;  
 CC cardiovascular; antianaemic; antileukemic; haemostatic; vulnery;  
 CC antidiabetic; osteopathic; dermatological; antiallergic; antisthmatic;  
 CC antidiabetic; cytostatic; neuroprotective; antidepressant; nootropic;  
 CC antiparkinsonian; and immunostimulant. The proteins and polynucleotides  
 CC encoding them can be used in gene therapy, antisense therapy and vaccine  
 CC production. The proteins and polynucleotides are useful for screening for  
 CC agonists or antagonists of a protein and for the treatment and diagnosis  
 CC of disorders associated with the activity of a protein e.g. inflammation,

CC rheumatoid arthritis, septic shock, pancreatitis, cardiac dysfunction,  
 CC neuropathology, cardiac anaphylaxis, viral, bacterial, HIV and fungal  
 CC infections, autoimmunity, genetic diseases, haematopoietic disorders,  
 CC anaemia, platelet disorders, thrombocytopaenia, wounds, burns, ulcers,  
 CC osteoporosis, severe combined immunodeficiency, eczema, allergic  
 CC rhinitis, asthma, diabetes, cancer, multiple sclerosis, depression,  
 CC Alzheimer's disease, Parkinson's disease, neurodegenerative and  
 CC neurological disorders.  
 XX  
 SO Sequence 152 AA;  
 QY Query Match 16.2%; Score 308.5; DB 22; Length 152;  
 Best Local Similarity 46.4%; Pred. No. 3.8e-18;  
 Matches 70; Conservative 21; Mismatches 43; Indels 17; Gaps 3;  
 DB 6 PNTTQFLMDHDDHDEP--DLKTGLYSKRAAKSDTSDDDFMEEGEGDGGSDG 259  
 202 PNTTQFLMDHDDHDEP--DLKTGLYSKRAAKSDTSDDDFMEEGEGDGGSDG 259  
 6 PNTTQFLMDHDDHDEPNDVPHGISHPSSGES-----EAGSDGRGRAH 51  
 QY 260 SEFLQRFSEYERYHTSLQNMKSQKELKEYLEKCLSRMEDENRRLRLSKRLGDD 319  
 52 GEFLQRFSEYERYHTSLQNMKSQKELKEYLEKCLSRMEDENRRLRLSKRLGDD 319  
 DB 52 GEFLQRFSEYERYHTSLQNMKSQKELKEYLEKCLSRMEDENRRLRLSKRLGDD 319  
 QY 320 AR-VRELEELRLRAENQLTENEHMQ 349  
 DB 112 CRQVELAEVORLRTENQRLROENQNMRE 142

RESULT 7  
 ABG15306  
 ID ABG15306 standard; Protein; 134 AA.  
 XX  
 AC ABG15306;  
 XX  
 DT 18-FEB-2002 (first entry)  
 XX  
 DE Novel human diagnostic protein #15297.  
 XX  
 KW Human; chromosome mapping; gene mapping; gene therapy; forensic;  
 KW food supplement; medical imaging; diagnostic; genetic disorder.  
 XX  
 OS Homo sapiens.  
 XX  
 PN WO200175067-A2.  
 XX  
 PD 11-OCT-2001.  
 XX  
 PF 30-MAR-2001; 2001WO-US08631.  
 XX  
 PR 31-MAR-2000; 2000US-0540217.  
 PR 23-AUG-2000; 2000US-0649167.  
 XX  
 PA (HYSE-) HYSEQ INC.  
 XX  
 PI Drmanac RT, Liu C, Tang YT;  
 XX  
 DR WPI; 2001-639362/73.  
 DR N-PSDB; AAS79493.  
 XX  
 PT New isolated polynucleotide and encoded polypeptides, useful in  
 PT diagnostics, forensics, gene mapping, identification of mutations  
 PT responsible for genetic disorders or other traits and to assess  
 PT biodiversity -  
 XX  
 Claim 20; SEQ ID No 45665; 103pp; English.

The invention relates to isolated polynucleotide (I) and  
 CC polypeptide (II) sequences. (I) is useful as hybridisation probes,  
 CC polymerase chain reaction (PCR) primers, oligomers, and for chromosome  
 CC and gene mapping, and in recombinant production of (II). The  
 CC polynucleotides are also used in diagnostics as expressed sequence tags  
 CC for identifying expressed genes. (I) is useful in gene therapy techniques  
 CC to restore normal activity of (II) or to treat disease states involving







QY 329 LDRLRARENILQLLT-----ENELHROQERAPL 354  
 DB 587 LDKMKKEHQVMAKARQVYAEERKQRAEL 616

RESULT 11  
 ABG20362  
 ID ABG20362 standard; Protein; 288 AA.  
 XX

AC ABG20362;

DT 13-FEB-2002 (first entry)

DE Novel human diagnostic protein #20353.

XX Human; chromosome mapping; gene mapping; gene therapy; forensic;  
 KM food supplement; medical imaging; diagnostic; genetic disorder.

XX Homo sapiens.

XX WO200175067-A2.

XX 11-OCT-2001.

PF 30-MAR-2001; 2001WO-US08631.

XX 31-MAR-2000; 2000US-0540217.

PR 23-AUG-2000; 2000US-0649167.

XX (HYSE-) HYSEQ INC.

PI Drmanac RT, Liu C, Tang YT;

DR WPI; 2001-639362/73.

XX N-PSDB; AAS84549.

PT New isolated polynucleotide and encoded polypeptides, useful in  
 PT diagnostics, forensic, gene mapping, identification of mutations  
 PT responsible for genetic disorders or other traits and to assess  
 PT biodiversity -

PS Claim 20; SEQ ID NO 50721; 103pp; English.

XX The invention relates to isolated polynucleotide (I) and  
 CC polypeptide (II) sequences. (I) is useful as hybridisation probes,  
 CC polymerase chain reaction (PCR) primers, oligomers, and for chromosome  
 CC and gene mapping, and in recombinant production of (II). The  
 CC polynucleotides are also used in diagnostics as expressed sequence tags  
 CC for identifying expressed genes. (I) is useful in gene therapy techniques  
 CC to restore normal activity of (II) or to treat disease states involving  
 CC (II). (II) is useful for generating antibodies against it, detecting or  
 CC quantitating a polypeptide in tissue, as molecular weight markers and as  
 CC a food supplement. (II) and its binding partners are useful in medical  
 CC imaging of sites expressing (II). (I) and (II) are useful for treating  
 CC disorders involving aberrant protein expression or biological activity.  
 CC The polypeptide and polynucleotide sequences have applications in  
 CC diagnostics, forensic, gene mapping, identification of mutations  
 CC responsible for genetic disorders or other traits to assess biodiversity  
 CC and to produce other types of data and products dependent on DNA and  
 CC amino acid sequences. ABG00010-ABG0377 represent novel human  
 CC diagnostic amino acid sequences of the invention.  
 CC Note: The sequence data for this patent did not appear in the printed  
 CC specification, but was obtained in electronic format directly from WIPO  
 CC at ftp.wipo.int/pub/published\_pct\_sequences.

XX Sequence 288 AA;

Query Match 7.5%; Score 143; DB 22; Length 288;

Best Local Similarity 27.5%; Pred. No. 0.00083;  
 Matches 69; Conservative 24; Mismatches 64; Indels 94; Gaps 14;

QY 50 RAFFPOL---GGRBG-----PEGGSLSQPPPLQTAQCPSSCLREGKGQNDGS 97

DB 9 RGVQLNSQAGPFGSGTAGDACPRAATGTL-----ALRLRA-----NSGERGQSLER 56  
 QY 98 SAGGDPPPAEVPTEAEFLAQPCHDSEASTLGAPAGE-----EEM-GQOQR----- 146  
 DB 57 S--GRAPPL-----LRGLGAGAGAGAGCGCGDWRGQGRPGRF 95  
 QY 147 -----QLGKKTKRRRPSKKRHWKPYKLTW-----EKKKPFDEK----- 181  
 DB 96 EIKKKLKTAKKKKKKKKKKEEEDKKLTQIQESQVTSNKKERRSKDEKLDKKSQAM 155  
 QY 182 QSLRASRIR-----AEWPAKGQPVAPYNTTQFLMDHDOEFPDLKGLYXSKAAASDPT 236  
 DB 156 EELKAREERKRNRTAELAKQPL--KTSVYSDDEEEEDD-----KSSEKSDRSRT 207  
 QY 237 SDDDFMEEGGE 247  
 DB 208 SSSDEEEKKEE 218

RESULT 12  
 ABB59642  
 ID ABB59642 standard; Protein; 1192 AA.  
 XX

AC ABB59642;

DT 26-MAR-2002 (first entry)

DE Drosophila melanogaster polypeptide SEQ ID NO 5718.

XX Drosophila; developmental biology; cell signalling; insecticide;

XX pharmaceutical.

OS Drosophila melanogaster.

XX WO200171042-A2.

XX 27-SEP-2001.

PF 23-MAR-2001; 2001WO-US09231.

XX 23-MAR-2000; 2000US-191637P.

PR 11-JUL-2000; 2000US-0614150.

XX (PEKE ) PE CORP NY.

XX Venter JC, Adams M, Li PWD, Myers EW;

DR WPI; 2001-656860/75.

XX N-PSDB; ABL03745.

PT New isolated nucleic acid detection reagent for detecting 1000 or more  
 PT genes from Drosophila and for elucidating cell signalling and cell-cell  
 PT interactions -

PS Disclosure; SEQ ID NO 5718; 21pp + Sequence Listing; English.

XX The invention relates to an isolated nucleic acid detection reagent  
 CC capable of detecting 1000 or more genes from Drosophila. The invention is  
 CC useful in developmental biology and in elucidating cell signalling and  
 CC cell-cell interactions in higher eukaryotes for the development of  
 CC insecticides, therapeutics and pharmaceutical drugs. The invention of  
 CC discloses genomic DNA sequences (ABL16176-ABL30511), expressed DNA  
 CC sequences (ABL01840-ABL16175) and the encoded proteins  
 CC (ABB57737-ABB72072).  
 CC The sequence data for this patent did not form part of the printed  
 CC specification, but was obtained in electronic format directly from WIPO  
 CC at ftp.wipo.int/pub/published\_pct\_sequences.

XX Sequence 1192 AA;

Query Match 7.4%; Score 141.5; DB 22; Length 1192;

Best Local Similarity 22.3%; Pred. No. 0.0072;



Matches 80; Conservative 57; Mismatches 138; Indels 83; Gaps 17;

QY 4 PFLSEYOHOPQTSNCTGAAGAAVOELNPERPGAEERVPEDSSWOSRAFPOLGRRPGEG 63  
 DB 252 PLNDELPEDESPAAITESAV--BELEKESRAAMDQVPESELOPEVGP---GEYQSES 306  
 QY 64 EG-----SLESQPPPLQTOACPESSCLREGEKGQNGDDSSAGGDFPPPAVEYPPPEAF 116  
 DB 307 DGEQAEETKPEIEAQP---EVEAQPEABAQPEAF-----PQLEVEPQPEVE 348  
 QY 117 LLAOPCHDSASKLGAAPAGGESEWQGOQOLKKHRRRPSKKKHMKVYVLTMEKK 176  
 DB 349 --SQPEVSEQPEVBAQPEVEPQSEVESQ-----PRAESISEPE--TQAEVE 390  
 QY 177 KPEKOSLRASRIAREMFAGQVAPVNTTQFLMDHDQEPDLTKGLYSKRAAASDDT 236  
 DB 391 AQPEVESLPBAESQPE--AESQ-----EPEVEPE---AEKISDNEVD 429  
 QY 237 SDDDFMEEGEGEDGSDGSDGSEFLQDPFSETTERYHTESIQNMSKQELIKEYLEK 296  
 DB 430 TEASLMEITVE--GIEDGLTAAMDNLVPEELAEASDQKQTE--LESEDDQSPVTEAIE-EQ 485  
 QY 297 CLSRMEDENRRLKLESKRLGDDARVLELELDRLAENLQ--LTENEIHRQOERAP 353  
 DB 486 AVPEIRQEKER---EPEQITLAD-----ETQDSAQPSNEPEVEIAPEQHTAEATAP 534

RESULT 13  
 AAY20666  
 ID AAY20666 standard; Protein; 557 AA.

XX AAY20666;  
 XX 22-JUL-1999 (first entry)  
 XX Human neurofilament-M wild type protein fragment 8.  
 DE Human neurofilament-M wild type protein fragment 8.  
 XX Human; beta-amyloid precursor protein; beta-APP; diagnosis; cancer;  
 KW frameshift mutation; age-related disease; neurodegenerative disorder;  
 KW Alzheimer's disease; Down's syndrome; myotonic dystrophy; neuronal;  
 KW Huntington's disease; multiple sclerosis; alcoholic liver disease;  
 KW diabetes mellitus type II; microtubule associated protein; Tau; Big Tau;  
 KW ubiquitin B; apolipoprotein E; MAP2; neurofilament-L; neurofilament-M;  
 KW neurofilament-F; presenilin I; presenilin II; cellular tumour antigen;  
 KW glial fibrillary acidic protein; GFAP; p53; semaphorin III; HUPF-1;  
 KW bcl-2; B-cell leukemia/lymphoma 2 proto-oncogene; HMGP-C; NSP-A;  
 KW high mobility group protein-C; neuroendocrine specific protein A.  
 XX Homo sapiens.  
 OS Homo sapiens.  
 XX W09845322-A2.  
 PN W09845322-A2.  
 XX 15-OCT-1998.  
 PD 15-OCT-1998.  
 XX 02-APR-1998; 98WO-1B00705.  
 PF 02-APR-1998; 98WO-1B00705.  
 XX 10-APR-1997; 97US-0043163.  
 PR 10-APR-1997; 97US-0043163.  
 XX (UYUT-) RIJXSUNITV UTRECHT.  
 PA (ROYA-) ROYAL NETHERLANDS ACAD ARTS & SCI.  
 PA (UYRO-) UNIV ROTTERDAM ERASMUS.  
 XX Burbach JPH, Grosveld FG, Van Leeuwen FW,  
 XX WPI, 1998-609901/51.  
 DR N-PSDB; AAX75759.  
 XX Diagnosing disease by detecting frameshift mutations in RNA or  
 PT corresponding protein mutations - used to diagnose cancer and  
 PT neurological diseases, particularly Alzheimer's disease, and also  
 PT for treatment and prevention with specific ribozymes or wild-type  
 PT RNA

PS Disclosure; Figure 8; 258bp; English.

XX This invention describes a novel method for the diagnosis of a disease  
 CC caused by, or associated with, an RNA molecule that has a frameshift  
 CC mutation. The method is used to diagnose age-related diseases, especially  
 CC cancer and a wide range of neurodegenerative disorders (e.g. Alzheimer's  
 CC disease, Down's syndrome, myotonic dystrophy, Huntington's disease, II  
 CC multiple sclerosis, alcoholic liver disease, diabetes mellitus type II  
 CC and many others listed) or susceptibility to these disorders. The method  
 CC allows a definitive diagnosis of Alzheimer's disease in living patients,  
 CC at an early stage. It is based on the observation that disease may be  
 CC caused by mutations in RNA rather than DNA. The invention describes the  
 CC use of neuronal system RNA molecules, specifically proteins including  
 CC beta-amyloid precursor protein (beta-APP), the microtubule associated  
 CC proteins Tau and Big Tau, ubiquitin B, apolipoprotein E, microtubule  
 CC associated protein 2 (MAP2), neurofilament-L, neurofilament-M,  
 CC neurofilament-F, presenilin I, presenilin II, glial fibrillary acidic  
 CC protein (GFAP), the cellular tumour antigen p53, B-cell leukemia/lymphoma  
 CC 2 (bcl-2) proto-oncogene, semaphorin III, HUPF-1, high mobility group  
 CC protein-C (HMGP-C) and neuroendocrine specific protein A.

XX Sequence 557 AA:

SO Query Match 7.2%; Score 138; DB 19; Length 557;  
 Best Local Similarity 22.5%; Pred. No. 0.0052;  
 Matches 80; Conservative 38; Mismatches 122; Indels 116; Gaps 17;

QY 31 ERPGAEERVPEDSSWOSRAFPOLGRRPGEG--SLESQPPPLQTOACPESSCLRGC 88  
 DB 4 ERPGGAGAPRAEKSPFCGL-----GPGAGAESFPSPPLPRCQSPPPR-- 53  
 QY 89 EKGQNGDDSSAGGDFPPPAVEPTPEAEILAPQCHDSKASKLGAAPAGGESEWQGOQOL 148  
 DB 54 -----PPPPRGGPWP-----GTSAALIN----- 71  
 QY 149 GKXKRRH-----RPSKKRRHW-----KPYUKLTWEKKKKFDEKOSLRASRIAREMRA 195  
 DB 72 GRRGRQNAVTATRPKASKSYTLDSLGNPSAYRRTETRSFSRVSQSSGFRSQSWS 131  
 QY 196 KQVPAPVNTTQFLMDHDQEPDLTKGLYSKX-----AAKSDDDSDDFMEEGEGEDG 251  
 DB 132 RSP-----STVSSSYKRSMLARLAYSSMBSAESSLDFSSSSILNG 177  
 QY 252 SGMGSDGSEFLQDPFSETTERYHTESIQNMSKQELIKEYLEKLSHMEDENNRLRE 311  
 DB 178 S-GPGSD-----YKLSRNEK---EQLOGLN--DRFAYIRK---VHLLDQONKEIEAF 222  
 QY 312 SKRLG-----GD--DARVLELELDRLAENLQ-----TENELHRQOER 351  
 DB 223 IQALRQKQASHAQLDGAYDQEIRELATLEVMNHEKAVQVLDSDHLEDDHRLKER 278

RESULT 14  
 AAY94461  
 ID AAY94461 standard; Protein; 530 AA.  
 XX AAY94461;  
 XX 11-SEP-2000 (first entry)  
 DT Human p75 protein.  
 XX Human p75 protein.  
 DE Human p75 protein.  
 XX p52; p75; transcription coactivator; ASF/SF2; gene therapy;  
 KW cancer; breast cancer.  
 XX Homo sapiens.  
 OS Homo sapiens.  
 XX Key Location/Qualifiers  
 FT Region 192..325  
 FT /label= Highly\_charged  
 XX W0200029578-A1.  
 XX

PD 25-MAY-2000.  
 XX  
 PF 10-NOV-1999; 99WO-US26792.  
 XX  
 PR 13-NOV-1998; 98US-0108248.  
 XX  
 PA (USSH ) US DEPT HEALTH & HUMAN SERVICES.  
 XX  
 PI Ge H;  
 XX  
 DR WPI: 2000-387791/33.  
 XX  
 DR N-PSDB; AAA27150.  
 PT Novel mRNA transcription coactivators p52 and p75 useful for treating  
 XX tumors and cancers comprising a specific amino acid sequence -  
 XX  
 PS Claim 6; Fig 3; 100pp; English.  
 XX  
 CC The N-terminal and two internal sequences were identified in a 75kDa  
 CC polypeptide that co-purified with a general transcription coactivator.  
 CC A degenerate oligonucleotide, based on the N-terminal sequence, was used  
 CC to screen a human cDNA library derived from an adenocarcinoma of the  
 CC cervix. From this the sequences of p52 and p75 were identified. The p52  
 CC and to a lesser extent p75 proteins were found to act as general  
 CC transcriptional coactivators. In addition p52 was found to enhance  
 CC ASF/SF2-mediated pre-mRNA splicing. The p52 and p75 polypeptides can be  
 CC used to treat disease caused by defects in transcription or in  
 CC ASF/SF2-mediated pre-mRNA splicing. These disease may also be treated by  
 CC gene therapy, using p52 and p75 polynucleotides. An example of a  
 CC treatable disease is cancer, especially breast cancer. The present  
 CC sequence is the human p75 protein, Genbank Accession No.  
 CC AAC97946.  
 XX  
 XX  
 SQ Sequence 530 AA;  
 Query Match 7.1%; Score 135.5; DB 21; Length 530;  
 Best Local Similarity 21.8%; Pred. No. 0.008;  
 Matches 79; Conservative 59; Mismatches 110; Indels 115; Gaps 18;  
 QY 7 SEYOHQOTSNGTGAALVOEELNPERPGAEBRVPEEDSRWQRAFPQ-----L 55  
 DB 93 SSQQAATKQSNMNSDVVEER-----ETSVSKEDTDHEKASNEDVTAKAVDITTPK 143  
 QY 56 GGRPG--PEGESLSQPPPLQTOACPESSCLREGEKQNGDSSAGDFPPPAEV---E 110  
 DB 144 AARRGRKRAEKQVETEEAGVTTATASVNLKVSPPKGR-----PATEVKIPK 192  
 QY 111 PTBEAELLAPCHDSEASKLGAAPAGEEEMGOORLGKKHRRRPSKKKHMKPYKYL 170  
 DB 193 PGRPRKRVKQPC-PSESDII-----TEEDSKKKKGQEGK-----OPKK----- 229  
 QY 171 TWEKKKPEDEKQSLRASRIAPAMPAKGPVAPYNTTQFLMDHDQEPDLKTG---LYSK 227  
 DB 230 ---QPKDEEGQK-----EDKPRKPEPKKGGKKEVESEK 260  
 QY 228 RA-AKSDTSDDDFMEEGGEEDGSDGCGSEFLORDFSETYER-----YHTESTIQ 280  
 DB 261 RKHLAKTGVSTSDSEEDGDDGEGEKRRKGG-----RNFGTAHRRNMLKQHEKEAAD 313  
 QY 281 NMSKQELIKYELLEKCLSRMEDENN---LRLESKRLGGDARVRELEL-DRLEAEN 336  
 DB 314 RKRKOE---EQMETEQ---QNKDEGKKPEVKVKEKRETSMDSLQRIHAETKNSLKIDN 367  
 QY 337 LQL 339  
 DB 368 LDV 370  
 RESULT 15  
 AAU74503  
 ID AAU74503 standard; Protein; 530 AA.  
 XX  
 AC AAU74503;

XX  
 DT 09-APR-2002 (first entry)  
 XX  
 DE Human nuclear transcription co-activator p75.  
 XX  
 KW Human; nuclear antigen; nuclear transcription co-activator; p75;  
 XX atopic dermatitis; asthma; interstitial cystitis; protein.  
 XX  
 OS Homo sapiens.  
 XX  
 FH Key Location/Qualifiers  
 FT Misc-difference 224 /note="Encoded by GTA"  
 XX  
 PN MO200190751-A1.  
 XX  
 PD 29-NOV-2001.  
 XX  
 PF 23-MAY-2001; 2001WO-US16639.  
 XX  
 PR 23-MAY-2000; 2000US-207414P.  
 XX  
 PA (SCRI ) SCRIPPS RES INST.  
 XX  
 PI Tan EM, Ochs RL, Chan EKL, Muro Y;  
 XX  
 DR WPI: 2002-122026/16.  
 DR N-PSDB; AAS20483.  
 XX  
 PT Screening a patient for atopic dermatitis, comprises determining the  
 PT presence of antibodies against nuclear transcription co-activator p75  
 PT in the sera of the patient -  
 XX  
 PS Disclosure; Page -: 22pp; English.  
 XX  
 CC The invention relates to a method of screening a patient for atopic  
 CC dermatitis, comprising determining the presence of antibodies against  
 CC nuclear transcription co-activator p75, in sera of the patient. The  
 CC method is used for diagnosing patients with atopic dermatitis, and for  
 CC detecting atopic dermatitis in patients suffering from other conditions  
 CC such as asthma or interstitial cystitis. This sequence represents the  
 CC human p75 protein.  
 CC  
 SQ Sequence 530 AA;  
 Query Match 7.1%; Score 135.5; DB 23; Length 530;  
 Best Local Similarity 21.8%; Pred. No. 0.008;  
 Matches 79; Conservative 59; Mismatches 110; Indels 115; Gaps 18;  
 QY 7 SEYOHQOTSNGTGAALVOEELNPERPGAEBRVPEEDSRWQRAFPQ-----L 55  
 DB 93 SSQQAATKQSNMNSDVVEER-----ETSVSKEDTDHEKASNEDVTAKAVDITTPK 143  
 QY 56 GGRPG--PEGESLSQPPPLQTOACPESSCLREGEKQNGDSSAGDFPPPAEV---E 110  
 DB 144 AARRGRKRAEKQVETEEAGVTTATASVNLKVSPPKGR-----PATEVKIPK 192  
 QY 111 PTBEAELLAPCHDSEASKLGAAPAGEEEMGOORLGKKHRRRPSKKKHMKPYKYL 170  
 DB 193 PGRPRKRVKQPC-PSESDII-----TEEDSKKKKGQEGK-----OPKK----- 229  
 QY 171 TWEKKKPEDEKQSLRASRIAPAMPAKGPVAPYNTTQFLMDHDQEPDLKTG---LYSK 227  
 DB 230 ---QPKDEEGQK-----EDKPRKPEPKKGGKKEVESEK 260  
 QY 228 RA-AKSDTSDDDFMEEGGEEDGSDGCGSEFLORDFSETYER-----YHTESTIQ 280  
 DB 261 RKHLAKTGVSTSDSEEDGDDGEGEKRRKGG-----RNFGTAHRRNMLKQHEKEAAD 313  
 QY 281 NMSKQELIKYELLEKCLSRMEDENN---LRLESKRLGGDARVRELEL-DRLEAEN 336  
 DB 314 RKRKOE---EQMETEQ---QNKDEGKKPEVKVKEKRETSMDSLQRIHAETKNSLKIDN 367

Fri Feb 6 07:22:04 2004

us-09-972-758a-2.rag

Page 11

QY	337	LQL	339
		:	
Db	368	LDV	370

Search completed: February 5, 2004, 13:28:25  
Job time : 83 secs

THIS PAGE BLANK (USPTO)

GenCore version 5.1.6  
Copyright (c) 1993 - 2004 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: February 5, 2004, 13:27:09 ; Search time 28 Seconds  
(without alignments)  
542.486 Million cell updates/sec

Title: US-09-972-758A-2  
Perfect score: 1910  
Sequence: 1 MAEPFLSEYQHPQTSNCTG.....LTENELHROOERAPLSKFGD 359

Scoring table: BLOSUM62  
Gapop 10.0 , Gapext 0.5

Searched: 328717 seqs, 42310858 residues

Total number of hits satisfying chosen parameters: 328717

Minimum DB seq length: 0  
Maximum DB seq length: 2000000000

Post-Processing: Minimum Match 0%  
Maximum Match 100%  
Listing first 45 summaries

Database : Issued Patents AA:\*  
1: /cgn2\_6/prodata/1/1aa/5A\_COMB.pep:\*  
2: /cgn2\_6/prodata/1/1aa/5B\_COMB.pep:\*  
3: /cgn2\_6/prodata/1/1aa/5A\_COMB.pep:\*  
4: /cgn2\_6/prodata/1/1aa/5B\_COMB.pep:\*  
5: /cgn2\_6/prodata/1/1aa/PCTUS\_COMB.pep:\*  
6: /cgn2\_6/prodata/1/1aa/backfill1.pep:\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	134.5	7.0	939	US-09-252-991A-19109	Sequence 19109, A
2	133	7.0	538	US-09-309-572-12	Sequence 12, App1
3	133	7.0	1312	US-09-554-572-26	Sequence 26, App1
4	133	7.0	1737	US-09-309-572-13	Sequence 13, App1
5	131	6.9	1898	US-08-056-200-94	Sequence 94, App1
6	131	6.9	1898	US-08-800-644-94	Sequence 94, App1
7	130	6.8	611	US-09-216-3938-81	Sequence 81, App1
8	125	6.5	754	US-09-214-564A-2	Sequence 2, App1
9	123.5	6.5	557	US-08-979-608A-5	Sequence 5, App1
10	123	6.4	1829	US-09-157-420-1	Sequence 1, App1
11	122	6.4	723	US-07-814-964-11	Sequence 11, App1
12	122	6.4	723	US-08-258-442-11	Sequence 11, App1
13	122	6.4	723	US-08-328-809-6	Sequence 6, App1
14	122	6.4	723	US-08-866-840-6	Sequence 6, App1
15	122	6.4	723	PCT-US92-11107-11	Sequence 11, App1
16	121	6.3	538	US-09-370-368-9	Sequence 9, App1
17	120	6.3	1780	US-08-769-309A-5	Sequence 5, App1
18	120	6.3	1780	US-08-994-570-5	Sequence 5, App1
19	118	6.2	820	US-09-252-991A-23346	Sequence 23346, A
20	118	6.2	830	US-09-562-737-39	Sequence 39, App1
21	116.5	6.1	369	US-09-252-991A-20245	Sequence 20245, A
22	116	6.1	530	US-08-979-608A-8	Sequence 8, App1
23	116	6.1	2972	US-09-579-181-2	Sequence 2, App1
24	116	6.1	3118	US-09-579-181-1	Sequence 1, App1
25	115.5	6.0	1596	US-08-578-277A-4	Sequence 4, App1
26	114.5	6.0	245	US-09-252-991A-29400	Sequence 29400, A
27	114.5	6.0	1958	US-07-945-283-2	Sequence 2, App1

28	114	6.0	683	6	5210183-3	Patent No. 5210183
29	113	5.9	380	4	US-09-252-991A-22319	Sequence 22319, A
30	113	5.9	8991	4	US-08-714-741-32	Sequence 32, App1
31	112.5	5.9	529	3	US-08-556-419-22	Sequence 22, App1
32	112.5	5.9	629	3	US-08-556-419-23	Sequence 23, App1
33	112.5	5.9	764	4	US-09-370-838-67	Sequence 67, App1
34	112	5.9	802	3	US-09-156-316-1	Sequence 1, App1
35	111	5.8	1162	2	US-08-728-323A-2	Sequence 2, App1
36	111	5.8	1162	4	US-09-298-568-2	Sequence 2, App1
37	111	5.8	1162	4	US-09-410-399-2	Sequence 2, App1
38	110	5.8	576	2	US-08-533-306A-2	Sequence 2, App1
39	110	5.8	576	2	US-08-742-923A-2	Sequence 2, App1
40	108.5	5.7	634	4	US-09-252-991A-24400	Sequence 24400, A
41	108.5	5.7	657	3	US-08-893-852A-3	Sequence 3, App1
42	108.5	5.7	657	3	US-08-821-818-3	Sequence 3, App1
43	108.5	5.7	657	4	US-09-052-753B-3	Sequence 3, App1
44	108	5.7	631	4	US-09-252-991A-20063	Sequence 20063, A
45	108	5.7	671	4	US-09-439-313-380	Sequence 380, App

ALIGNMENTS

RESULT 1									
US-09-252-991A-19109									
Sequence 19109, Application US/09252991A									
Patent No. 6551795									
GENERAL INFORMATION:									
APPLICANT: Marc J. Rubenfield et al.									
TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO PSEUDOMONAS									
FILE REFERENCE: 107196.136									
CURRENT APPLICATION NUMBER: US 09/252,991A									
PRIOR FILING DATE: 1999-02-18									
PRIOR APPLICATION NUMBER: US 60/074,788									
PRIOR FILING DATE: 1998-02-18									
PRIOR APPLICATION NUMBER: US 60/094,190									
PRIOR FILING DATE: 1998-07-27									
NUMBER OF SEQ ID NOS: 33142									
SEQ ID NO 19109									
LENGTH: 939									
TYPE: PRT									
ORGANISM: Pseudomonas aeruginosa									
US-09-252-991A-19109									
Query Match									
Best/Local Similarity 7.0%; Score 134.5; DB 4; Length 939;									
Matches 88; Conservative 44; Mismatches 146; Indels 145; Gaps 17;									
QY	12	OPOTNCTGAAGVQELNPERPPGABERYVEEDSRMQSAPFQLOG-----RPGPEG	63						
QY	47	QPEADRTA-----RLRPRRPPGSRRRROPARRLQSGHALPVGCVPLVPAADOPRA	101						
DB	64	EGSLSEOP-PPLOTQACPESSCLARG-----EKXGNDSSAGDPP-PPA	107						
QY	102	YPAQGRPAQPLAADPAGSSVARSAGTRDLRLMRRRRRAGSDASRSRSHHPRA	161						
DB	108	EVEIPPEAEV-----LAQCHDEA-----SKGAPA-----	134						
QY	162	RLYPRADRLPPSGADGRHEHPRRRSSGKLAPARPPDPAARCDVSSQLAAAGSHPA	221						
DB	135	-----AGGEEMWGOQORO-LGKKKRRRPPSKK-----RHMKPY	168						
QY	222	QRRSSRACRRAPRPPDLRPSAGIGDSPAQGGARCGARCRSAMSRRDVQKKKY	281						
DB	169	KLTWEKKKKFDEKQSLRASRIAEWFAKQGPVAPYNTTQFLMDHDQEBPDILKGLYSKR	228						
QY	289	LENIGQERLQRRMDARIDLRL-----RGVLRSS	310						
DB	229	AAAKDDTSDDDFME-----GGEEDGSDMGDGSFLODRDSEYERHTESLQNM	282						
QY	311	LAAGSDKAVDQCKMELREILRRDDMDAGLSGL-----IPRLKAVLDSEQRQORTQONI	366						

QY 283 -SKOELIKEXLEYLEKCLSMEDENNRLRLSKRLGSD-DARVE-----LELELDRLRA 334  
 DB 367 DALGEIAQQLAL-----DLPRRLRKLKQFARDIERARQSRNEIPILSELRLQR 418  
 QY 335 ENL 337  
 DB 419 QAL 421

## RESULT 2

US-09-309-572-12  
 ; Sequence 12, Application US/09309572  
 ; Patent No. 6440730  
 ; GENERAL INFORMATION:  
 ; APPLICANT: Heinrich-Pette-Institut  
 ; TITLE OF INVENTION: Retroviral hybrid vectors pseudotyped with LCMV  
 ; FILE REFERENCE: P50489  
 ; CURRENT APPLICATION NUMBER: US/09/309,572  
 ; CURRENT FILING DATE: 1999-05-11  
 ; EARLIER APPLICATION NUMBER: DE 198 56 463  
 ; EARLIER FILING DATE: 1998-11-26  
 ; NUMBER OF SEQ ID NOS: 24  
 ; SOFTWARE: Patentin Ver. 2.0  
 ; SEQ ID NO 12  
 ; LENGTH: 538  
 ; TYPE: PR1  
 ; ORGANISM: Moloney murine leukemia virus  
 ; FEATURE:  
 ; OTHER INFORMATION: gag protein  
 US-09-309-572-12

Query Match 7.0%; Score 133; DB 4; Length 538;  
 Best Local Similarity 20.9%; Pred. No. 0.0024;

Matches 89; Conservative 40; Mismatches 142; Indels 154; Gaps 19;

QY 30 PERPPGA-----EERVPEEDSRWQSRAPFOLGGRPE--GEQS-----LESOPPL 74  
 DB 107 PLPPSAPSLPLEPPRSTPPRSSLYPALT-PSLGAKPKQVLSGGLIDLTEDPPPY 165  
 QY 75 QTOACPESSCLREGEKQNGDDSSAGDPPPAVEPTPEALL-----AQCHDSASKL 130  
 DB 166 RDRPPPSD--RDG-----NGEATPAGEAP-----DPSPWASRLGRREPPVADSTTSA 214  
 QY 131 GAPAAAGEEEMWQOQROLGKKHRRRPSKKRHWKPYKLTWEKKKFKPEKQSLRASRIR 190  
 DB 215 FLRAAGN-----GQLQYWPSSSDLYNWK-----NNPSFSDPKLTALLIE 257  
 QY 191 AEMFAKQPVAPYNTTQFLMDHDQEPDLKTG-----LYSKRAAKSDD----- 235  
 DB 258 SVLI-----THQPTWDDCCQLLGLTLTGEEKQVLLFARKAVRGDDGRPTQLPNE 307  
 QY 236 -----TSDDDFMEEGE-----DGSGMGDGE 261  
 DB 308 VDAAPLERPDWDYTTQAGRNHLVHYRQLLAGLQNAGRSPTNLAKVKITQSPNESPSA 367  
 QY 262 FLORDSEYERY-----HTESLONMSKOELIKE 290  
 DB 368 FLER-LKEAYRRKTYTPDPEPQETVMSFIWQSPADIGKLERLEDLKNKTLGLDVR- 425  
 QY 291 YLELEKCLSMEDENNRLRLSKRLGSDARV-RELELELDRLRAENLQLTENELHROQ 349  
 DB 426 --EAKTIFNKRTPEER-----EERIRRETEKEERARRTEDEQEKERDRRRHR 472  
 QY 350 ERAPL 354  
 DB 473 EMSKL 477

## RESULT 3

US-09-554-572-26  
 ; Sequence 26, Application US/09554572  
 ; Patent No. 6573091  
 ; GENERAL INFORMATION:

; APPLICANT: NATURE TECHNOLOGY, INC.  
 ; TITLE OF INVENTION: CHIMERIC VIRAL PACKAGING SIGNAL WITHOUT GAG GENE  
 ; FILE REFERENCE: 228,000,302,01  
 ; CURRENT APPLICATION NUMBER: US/09/554,572  
 ; CURRENT FILING DATE: 2000-09-18  
 ; NUMBER OF SEQ ID NOS: 26  
 ; SOFTWARE: Patentin Ver. 2.0  
 ; SEQ ID NO 26  
 ; LENGTH: 1312  
 ; TYPE: PR1  
 ; ORGANISM: Murine leukemia virus  
 US-09-554-572-26

Query Match 7.0%; Score 133; DB 4; Length 1312;  
 Best Local Similarity 20.9%; Pred. No. 0.0078;

Matches 89; Conservative 40; Mismatches 142; Indels 154; Gaps 19;

QY 30 PERPPGA-----EERVPEEDSRWQSRAPFOLGGRPE--GEQS-----LESOPPL 74  
 DB 217 PLPPSAPSLPLEPPRSTPPRSSLYPALT-PSLGAKPKQVLSGGLIDLTEDPPPY 275  
 QY 75 QTOACPESSCLREGEKQNGDDSSAGDPPPAVEPTPEALL-----AQCHDSASKL 130  
 DB 276 RDRPPPSD--RDG-----NGEATPAGEAP-----DPSPWASRLGRREPPVADSTTSA 324  
 QY 131 GAPAAAGEEEMWQOQROLGKKHRRRPSKKRHWKPYKLTWEKKKFKPEKQSLRASRIR 190  
 DB 325 FLRAAGN-----GQLQYWPSSSDLYNWK-----NNPSFSDPKLTALLIE 367  
 QY 191 AEMFAKQPVAPYNTTQFLMDHDQEPDLKTG-----LYSKRAAKSDD----- 235  
 DB 368 SVLI-----THQPTWDDCCQLLGLTLTGEEKQVLLFARKAVRGDDGRPTQLPNE 417  
 QY 236 -----TSDDDFMEEGE-----DGSGMGDGE 261  
 DB 418 VDAAPLERPDWDYTTQAGRNHLVHYRQLLAGLQNAGRSPTNLAKVKITQSPNESPSA 477  
 QY 262 FLORDSEYERY-----HTESLONMSKOELIKE 290  
 DB 478 FLER-LKEAYRRKTYTPDPEPQETVMSFIWQSPADIGKLERLEDLKNKTLGLDVR- 535  
 QY 291 YLELEKCLSMEDENNRLRLSKRLGSDARV-RELELELDRLRAENLQLTENELHROQ 349  
 DB 536 --EAKTIFNKRTPEER-----EERIRRETEKEERARRTEDEQEKERDRRRHR 582  
 QY 350 ERAPL 354  
 DB 583 EMSKL 587

## RESULT 4

US-09-309-572-13  
 ; Sequence 13, Application US/09309572  
 ; Patent No. 6440730  
 ; GENERAL INFORMATION:  
 ; APPLICANT: Heinrich-Pette-Institut  
 ; TITLE OF INVENTION: Retroviral hybrid vectors pseudotyped with LCMV  
 ; FILE REFERENCE: P50489  
 ; CURRENT APPLICATION NUMBER: US/09/309,572  
 ; CURRENT FILING DATE: 1999-05-11  
 ; EARLIER APPLICATION NUMBER: DE 198 56 463  
 ; EARLIER FILING DATE: 1998-11-26  
 ; NUMBER OF SEQ ID NOS: 24  
 ; SOFTWARE: Patentin Ver. 2.0  
 ; SEQ ID NO 13  
 ; LENGTH: 1737  
 ; TYPE: PR1  
 ; ORGANISM: Moloney murine leukemia virus  
 ; FEATURE:  
 ; OTHER INFORMATION: gag-pol protein  
 US-09-309-572-13

Query Match 7.0%; Score 133; DB 4; Length 1737;  
 Best Local Similarity 20.9%; Pred. No. 0.011;  
 Matches 89; Conservative 40; Mismatches 142; Indels 154; Gaps 19;

30 PERPPA-----ERRVEEDSRWQSAFPOLGSGPGE-----GEES-----LESGPPPL 74  
 107 PPLPPAPPLPPLPPSTPPSSLYPALT-PSLGAKPKQVUSDSGLDILDTEDPPPY 165  
 75 OTQACPESSCLREGEKONGDDSSAGDPPPAVEPTPEAEIL-----AQCHDSEASKL 130  
 166 RDRPPPSD-RDG-----NGENTPAGAP-----DPSMASLRGRREPPVADSTTQA 214  
 131 GAPACGSEEWGQOQOLGKKRRRRPSKKRRHWKYLTWEKKKFEKOSLRASRIR 190  
 215 PLRAGGN-----GOLQYWPSSSDLYNMK-----NNNPSFSDPKLTALIE 257  
 191 AEMFAGQVAPYNTTQFLMDHDHGEPLDKT-----LYSRAAKSDD----- 235  
 258 SVLI-----THQPTWDDCQQLLGLTLTGEEKQVLLKAKAVRDDGPTQLPNE 307  
 236 -----TSDDFMEEGGE-----DGSDGMGGDGE 261  
 308 VDAAPLEPDDYTTQAGRNHLVHYRQLLAGLQAGSPYTLAKVGTGQPNESPA 367  
 262 FLORDESEYERY-----HTESLQNSKQELIKE 290  
 368 FLER-LKAYRYRTPYDPEDPQETVNSFWQAPDGRKLERLEDLKNKTLGDLVR- 425  
 291 YLELECLSRMEDENRRLKESKRLGDDARV-RELLELDRLRAANLQLTENELHQO 349  
 426 --EAEKTFPKRTEPBER-----BERIRRETEKEKERRRTEDEKERRRRR 472  
 350 ERAPL 354  
 473 EMSKL 477

RESULT 5  
 US-08-056-200-94  
 ; Sequence 94, Application US/08056200  
 ; Patent No. 5616500  
 ; GENERAL INFORMATION:  
 ; APPLICANT: Steinert, Peter M.  
 ; APPLICANT: Lee, Seung-Chul  
 ; APPLICANT: Kim, In-Gyu  
 ; APPLICANT: Chung, Soo-Il  
 ; APPLICANT: Park, Sang-Chul  
 ; TITLE OF INVENTION: Trichohyalin and Transglutaminase-3 and  
 ; TITLE OF INVENTION: Methods of Using Same  
 ; NUMBER OF SEQUENCES: 117  
 ; CORRESPONDENCE ADDRESS:  
 ; ADDRESSEE: Knobbe, Martens, Olson & Bear  
 ; STREET: 620 Newport Center Drive, Sixteenth Floor  
 ; CITY: Newport Beach  
 ; STATE: CA  
 ; COUNTRY: U.S.A.  
 ; ZIP: 92660  
 ; COMPUTER READABLE FORM:  
 ; MEDIUM TYPE: Floppy disk  
 ; COMPUTER: IBM PC compatible  
 ; OPERATING SYSTEM: PC-DOS/MS-DOS  
 ; SOFTWARE: Patentin Release #1.0, Version #1.25  
 ; CURRENT APPLICATION DATA:  
 ; APPLICATION NUMBER: US/08/056,200  
 ; FILING DATE: 30-APR-1993  
 ; CLASSIFICATION: 435  
 ; ATTORNEY/AGENT INFORMATION:  
 ; NAME: Fedrick, Michael F.  
 ; REGISTRATION NUMBER: 36,799  
 ; REFERENCE/DOCKET NUMBER: NIH054.001A  
 ; TELECOMMUNICATION INFORMATION:  
 ; TELEPHONE: (714) 760-0404  
 ; TELEFAX: (714) 760-9502

INFORMATION FOR SEQ ID NO: 94:  
 ; SEQUENCE CHARACTERISTICS:  
 ; LENGTH: 1898 amino acids  
 ; TYPE: amino acid  
 ; TOPOLOGY: linear  
 ; MOLECULE TYPE: protein  
 ; US-08-056-200-94

Query Match 6.9%; Score 131; DB 1; Length 1898;  
 Best Local Similarity 22.4%; Pred. No. 0.019;  
 Matches 79; Conservative 54; Mismatches 146; Indels 74; Gaps 13;

31 ERPPAER-----VPEEDSRWQSAFPOLGSGPGEBSLSQPPPLTQACPESS 83  
 813 ERLAERKQNGEQRFLEEEKEQR-----GRRREREKELQFLREELQERRRQ 865  
 84 CLREKQNGDSSAGDPPPAVEPTPEAE-----LLAQCHDSEASKLAPAPAG 136  
 866 QLOEEDGLQEOERRROQRDDQKRWQLEERKRRRHTLYAKPALQELRKQQLQOE 925  
 137 GEEWGQOQOLGKKRRRRPSKKRRHWKYLTWEKKKFEKOSLRASIRAEWFAK 196  
 926 EEEELQREER-----KRRQEQRYREELQOEEQLREERKRRRRERQRYRK 979  
 197 GQPVAPYNTTQFLMDHDQ--EERDLKGLYSKRAAKSDTSDDFMEEGSEEDGSD 253  
 980 DK-----LQKEEQLQEEPEKR-----RQREKRYREBELQOE----- 1016  
 254 GNGGSEFLQDFSEYER-----YHTESLQNSKQELIKEYLELECLSRMED--EN 305  
 1017 -----EQLLEERKRRROEMERYRKQELQOEEQLLEERKRLQERRQYRE 1070  
 306 NLRLESKRLGGDDARV--ELELEL--DLRAENLQLTEN-ELHQQR 351  
 1071 ERLQOEEQLGEEFRTRRQELERQYRKKEBELQOEEQLLEERPKRRRQR 1123

RESULT 6  
 US-08-800-644-94  
 ; Sequence 94, Application US/08800644  
 ; Patent No. 5958752  
 ; GENERAL INFORMATION:  
 ; APPLICANT: Steinert, Peter M.  
 ; APPLICANT: Lee, Seung-Chul  
 ; APPLICANT: Kim, In-Gyu  
 ; APPLICANT: Chung, Soo-Il  
 ; APPLICANT: Park, Sang-Chul  
 ; TITLE OF INVENTION: Trichohyalin and Transglutaminase-3 and  
 ; TITLE OF INVENTION: Methods of Using Same  
 ; NUMBER OF SEQUENCES: 117  
 ; CORRESPONDENCE ADDRESS:  
 ; ADDRESSEE: Knobbe, Martens, Olson & Bear  
 ; STREET: 620 Newport Center Drive, Sixteenth Floor  
 ; CITY: Newport Beach  
 ; STATE: CA  
 ; COUNTRY: U.S.A.  
 ; ZIP: 92660  
 ; COMPUTER READABLE FORM:  
 ; MEDIUM TYPE: Floppy disk  
 ; COMPUTER: IBM PC compatible  
 ; OPERATING SYSTEM: PC-DOS/MS-DOS  
 ; SOFTWARE: Patentin Release #1.0, Version #1.25  
 ; CURRENT APPLICATION DATA:  
 ; APPLICATION NUMBER: US/08/800,644  
 ; FILING DATE: 14-FEB-1997  
 ; CLASSIFICATION: 424  
 ; PRIOR APPLICATION DATA:  
 ; APPLICATION NUMBER: US 08/056,200  
 ; FILING DATE: 30-APR-1993  
 ; ATTORNEY/AGENT INFORMATION:  
 ; NAME: Fedrick, Michael F.  
 ; REGISTRATION NUMBER: 36,799  
 ; REFERENCE/DOCKET NUMBER: NIH054.001A

TELECOMMUNICATION INFORMATION:  
 TELEPHONE: (714) 760-0404  
 TELEFAX: (714) 760-9502  
 INFORMATION FOR SEQ ID NO: 94:  
 SEQUENCE CHARACTERISTICS:  
 LENGTH: 1898 amino acids  
 TYPE: amino acid  
 TOPOLOGY: linear  
 MOLECULE TYPE: protein  
 US-08-800-644-94

Query Match 6.9%; Score 131; DB 2; Length 1898;  
 Best Local Similarity 22.4%; Pred. No. 0.019;  
 Matches 79; Conservative 54; Mismatches 146; Indels 74; Gaps 13;

QY 31 ERPPGAER-----VPEEDSRWQSRAPFOLGSPPEGEGLSQQPPLQTOACPS 83  
 DB 813 ERQLRAEERQORERFPPEEKEQR-----GRQRERKEQLPPEEEOQLORRRAQ 865  
 QY 84 CLAREGKQNGDDSSAGDFPPPAVEPTPEAE-----LLAOPCHDSEASKLAPAG 136  
 DB 866 QLOEEDBDLQEDDERRRQORQDQKRWQLEEEERKRHTLTAKPALQEQLRKEQLQE 925  
 QY 137 GEEMGQOQOLGKKRRRPSKKRHKRYKLTWEKKKPKPKQSLRASRIAEKFAK 196  
 DB 926 EEEELORERE-----KRRROEOERYREEOQLQOEEOQLREERKRROERORYK 979  
 QY 197 GQVAVPYNTQFLMDHDQ---EEDPLKTGLYSKAAAKSDDTSDDFMEGGEEDGSD 253  
 DB 980 DKR-----LQKEEOQLGSEPEKR-----RRQREKKYEEELQOE----- 1016  
 QY 254 GMGDGSSEFLQDPSFETYEY-----YHTESLQMSKQELIKYLELEKLSMED--EN 305  
 DB 1017 -----EQQLRERERKRROERYKKEQLQOEEOQLREERKRROERORYRE 1070  
 QY 306 NRRLSEKRLGSGDARVR---ELEEL---DRRAEYLQLTEN-ELHROER 351  
 DB 1071 EELQOEEOQLGERETRRQERLEORYKEEELQOEEOQLREPEKRRQER 1123

RESULT 7  
 US-09-216-393B-81  
 ; Sequence 81; Application US/09216393B  
 ; Patent No. 6514694  
 ; GENERAL INFORMATION:  
 ; APPLICANT: Milhausen, Michael James  
 ; TITLE OF INVENTION: TOXOPLASMA GONDII PROTEINS, NUCLEIC ACID MOLECULES, AND USES THEREOF  
 ; FILE REFERENCE: TX-1-C2  
 ; CURRENT APPLICATION NUMBER: US/09/216, 393B  
 ; CURRENT FILING DATE: 1998-12-18  
 ; PRIOR APPLICATION NUMBER: 08/994, 825  
 ; PRIOR FILING DATE: 1997-12-19  
 ; NUMBER OF SEQ ID NOS: 366  
 ; SOFTWARE: PatentIn version 3.1  
 ; SEQ ID NO 81  
 ; LENGTH: 611  
 ; TYPE: PRT  
 ; ORGANISM: Toxoplasma gondii  
 ; US-09-216-393B-81

Query Match 6.8%; Score 130; DB 4; Length 611;  
 Best Local Similarity 22.2%; Pred. No. 0.0052;  
 Matches 82; Conservative 58; Mismatches 155; Indels 74; Gaps 15;

QY 13 POTSNTGAAVQELN-PPRPGAERVP-----EEDSRWQSRAPF-----QCGRRPPE 62  
 DB 189 PRSSKSDVCCSPQARUSLPQOSJGSSPSPISVTNDVYALFDSSASPLAHGELSSLPAGAV 248  
 QY 63 GEGSLSSQPP-----PLQTOACPSSECLAREGKQNGDDSSAGDFPPPAVEPTPEAE 118  
 DB 249 SAERLLTAPAEIAGPSASACLSVSC-----GPGMSPTADTTR- 287  
 QY 119 AOPCHDSEASKLAPAGGE-----EEMGQOQOLGKKRRRPSKKRHKRYKLTWE 174

DB 288 ----HDAERERRRAEEEXERERQEEERERRRVEEKKERQEEERE-----RRVEE 339  
 QY 175 K--KKDEKQSLRA-----SRIRAEWFAKQVAPYNTQFLMDHDQEPDLKTGLYSK 227  
 DB 340 KARQREDEDERRRRVEEKAQOREEERERRRVEEKAQ-----KEEERERRRVEEK 395  
 QY 228 RAAKSDDTSDDFMEB-----GGEEDGSGMGQDGSSEFLQDPSFETYE--RYHTESLQ 281  
 DB 396 ARQREERERRRRVVEEKAQOREEERERRRVEEKAQOREEERERRRVEEKAQ 455  
 QY 282 MSKQELIKYLELEKLSMEDENNRRLRESKRLGGDDARVLELELRLAENILQLT 341  
 DB 456 REEERERRRRVVEEKAQOREEERERRRVEEKAQ-----ERERQEEERERRRVEE 505  
 QY 342 ENELHROE 350  
 DB 506 EKERERQEE 514

RESULT 8  
 US-09-214-564A-2  
 ; Sequence 2; Application US/09214564A  
 ; Patent No. 6150315  
 ; GENERAL INFORMATION:  
 ; APPLICANT: Sharp, Phillip A.  
 ; APPLICANT: Zhou, Qiang  
 ; TITLE OF INVENTION: TAT-SF: Cofactor For Stimulation Of Transcriptional  
 ; TITLE OF INVENTION: Elongation By HIV-1 TAT  
 ; FILE REFERENCE: M0656/7042  
 ; CURRENT APPLICATION NUMBER: US/09/214, 564A  
 ; CURRENT FILING DATE: 1999-08-18  
 ; PRIOR APPLICATION NUMBER: US 60/021,218  
 ; PRIOR FILING DATE: 1996-07-03  
 ; PRIOR APPLICATION NUMBER: US 60/033,152  
 ; PRIOR FILING DATE: 1996-12-13  
 ; PRIOR APPLICATION NUMBER: PCT/US97/11713  
 ; PRIOR FILING DATE: 1997-07-03  
 ; NUMBER OF SEQ ID NOS: 5  
 ; SOFTWARE: FastSeq for Windows Version 3.0  
 ; SEQ ID NO 2  
 ; LENGTH: 754  
 ; TYPE: PRT  
 ; ORGANISM: Homo sapiens  
 ; US-09-214-564A-2

Query Match 6.5%; Score 125; DB 3; Length 754;  
 Best Local Similarity 21.5%; Pred. No. 0.016;  
 Matches 82; Conservative 50; Mismatches 166; Indels 84; Gaps 17;

QY 9 YOHOPOTSNTGAAVQELNPPRPGAERVPPEEDSRWQSRAPFOLG--RPGEDEGS 66  
 DB 401 FSHSTSGMA-----QETATGMAFEEPIDEKTEK---TEDGEFEGASENNA 448  
 QY 67 LESQPPPLQTOACP---ESSCLAREGKQNGDDSSAGDFPPPAVEPTPEAE---LL 118  
 DB 449 KESSPEKEAEBCPEKSESEGCPRKGFEGSCQKSESEBGN-FVRGSEEDSPKTEKKTLL 507  
 QY 119 AOPCHDSEASKLAPAGGEEMGQOQOLGKKRRRPSKKRHKRYKLTWEKKKF 178  
 DB 508 KNDCEBNGLAKESEDINKSE-----EVEGRTKSEEDDSK-----ESDEDC 551  
 QY 179 DEKQSLRASRIAEWFAKQVAPYNTQFLMDHDQ---EEDPLKTGLYSKRAAKSDPT 236  
 DB 552 SKQSEDESERFE-----ENGLEKDLDEBSEKELHENVLDKE--LEENDS 596  
 QY 237 SDDDFMEBEGEEDGSGMGQDGSSEFLQDPSFETYE-----TYER-YHTESLQMSKQ 285  
 DB 597 ENSEF-----EEDGSEKVLDEEGSE---REFDEDSDEKEEEDTYEKVFDDESDEKDE 648  
 QY 286 ELIKYLELEKLSMEDENNRRLRESKRLGDDARVLELELRLAENILQLT 342  
 DB 649 VADKGLAADKKAEBGADKTLFEESDDKEDDADGKEVADKTLFEDDSNEKLTFDE 708



QY 343 NE-----LHRQGERAPLSKFG 358  
 Db 709 EEDSSSEKLFDDSDERGTGFGFG 730

RESULT 9  
 US-08-979-608A-5  
 ; Sequence: 5, Application US/08979608A  
 ; Patent No. 6355451  
 ; GENERAL INFORMATION:  
 ; APPLICANT: Lees, Ann M.  
 ; Lees, Robert S.  
 ; Law, Simon W.  
 ; Arjona, Anibal A.  
 ; TITLE OF INVENTION: NOVEL LOW DENSITY LIPOPROTEIN  
 ; BINDING PROTEINS AND THEIR USES IN DIAGNOSING AND  
 ; TREATING ATHEROSCLEROSIS

NUMBER OF SEQUENCES: 42  
 CORRESPONDENCE ADDRESS:  
 ADDRESSEE: Fish & Richardson P.C.  
 STREET: 225 Franklin Street  
 CITY: Boston  
 STATE: MA  
 COUNTRY: USA  
 ZIP: 02110-2804  
 COMPUTER READABLE FORM:  
 MEDIUM TYPE: Diskette  
 OPERATING SYSTEM: DOS  
 SOFTWARE: FastSeq for Windows Version 2.0  
 CURRENT APPLICATION DATA:  
 APPLICATION NUMBER: US/08/979,608A  
 FILING DATE: 26-NOV-1997  
 PRIOR APPLICATION DATA:  
 APPLICATION NUMBER: US 60/048,547  
 FILING DATE: 03-JUN-1997  
 APPLICATION NUMBER: US 60/031,930  
 FILING DATE: 27-NOV-1996  
 ATTORNEY/AGENT INFORMATION:  
 NAME: Myers, Louis  
 REGISTRATION NUMBER: 35,965  
 REFERENCE/DOCKET NUMBER: 10797-002001 (formerly 3983/59818)  
 TELECOMMUNICATION INFORMATION:  
 TELEPHONE: 617/542-5070  
 TELEFAX: 617/542-8906  
 INFORMATION FOR SEQ ID NO: 5:  
 SEQUENCE CHARACTERISTICS:  
 LENGTH: 557 amino acids  
 TYPE: amino acid  
 TOPOLOGY: linear  
 MOLECULE TYPE: protein  
 FRAGMENT TYPE: internal  
 SEQUENCE DESCRIPTION: SEQ ID NO: 5:  
 US-08-979-608A-5

Query Match 6.5%; Score 123.5; DB 4; Length 557;  
 Best Local Similarity 22.6%; Pred. No. 0.017;  
 Matches 91; Conservative 53; Mismatches 123; Indels 135; Gaps 20;

QY 20 GAAVQELNPERPPGAERVPE-----EDSRMOSRAPFOLGRRPPEGGESLSQ 70  
 Db 9 GAA---KQNPKSPGPPGAGAGAGQRRPAPARAAGASGAPR--PEGQAQTAQ 63  
 QY 71 PEPV-----QTQACPESSCLREGEKQNGDSSAGDPFPPAEVPTPEAEILAP 121  
 Db 64 PGLALCVSEBELRQLEIDILSTYCV-DNNQCAPGEDGVQ-----EP-PEPE----- 107  
 QY 122 CHNSEAKL-----GAPAGGEEHNGQQ-----RQLGKKHRRRPSKKR 162  
 Db 108 --DNEKRAVAVARNGEPGTPVNGEKETSKAPGTGEIRTSDEVDGRHRRPQEKKA 165  
 QY 163 HW-----KPYXKLWEEKKKFDEKQSLRASRIRAEWFAKQGPVAPYNTT 206

Db 166 KGLGKEITLMTQTLNTLSTPEEKALALCKYAELEBHNSSOKMQLQKQ-----S 218  
 QY 207 QFLMDHDQDEBPDLTKGLYSKRAAKSDTS-----DDFMWEGGEDSGDMGSD 258  
 Db 219 QLV-----QEKDHLRGHSHKAILARSKLESICRELQRRNRSLKEG----- 259  
 QY 259 GSEFLQDFSEFYERYHTESLQNMKSQELIKYLELEKCLSMEDENNRLRLSKRLGSD 318  
 Db 260 ---VQRAEEEEKREKVTSHFQMTLNDI---QLQME---QHNRNSKTLRQENNELA-- 306  
 QY 319 DARVELEHLDRLRAENL-----QLTNEHLHQOE 350  
 Db 307 -ERLKL-TEOYELREEHIDVFKHKDLOQLQVDAKLOQAOE 346

RESULT 10  
 US-09-157-420-1  
 ; Sequence 1, Application US/09157420  
 ; Patent No. 6180760  
 ; GENERAL INFORMATION:  
 ; APPLICANT: TAKAI, Yoshimi  
 ; APPLICANT: NAKAMISHI, Hiroyuki  
 ; APPLICANT: MANDAI, Kenji  
 ; APPLICANT: WADA, Manabu  
 ; APPLICANT: OBAISHI, Hiroshi  
 ; TITLE OF INVENTION: ACTIN FILAMENT-BINDING PROTEIN "L-ARADIN"  
 ; FILE REFERENCE: 98-1042\*/LC(WMC)/653  
 ; CURRENT APPLICATION NUMBER: US/09/157,420  
 ; CURRENT FILING DATE: 1998-09-21  
 ; NUMBER OF SEQ ID NOS: 1  
 ; SOFTWARE: Patentin Ver. 2.0  
 ; SEQ ID NO 1  
 ; LENGTH: 1829  
 ; TYPE: PRT  
 ; ORGANISM: rat  
 ; US-09-157-420-1

Query Match 6.4%; Score 123; DB 3; Length 1829;  
 Best Local Similarity 19.9%; Pred. No. 0.087;  
 Matches 83; Conservative 59; Mismatches 157; Indels 116; Gaps 15;

QY 9 YHQPQTSNCTGAAY-----QELNPERPPGAERVPE-----EDSRMOSRAPF 53  
 Db 1259 YEEKPMHTESDHASIAIGVTRSGEELREKVVQLERHVRVSGMDKCDSDWIMQS-- 1316  
 QY 54 QLGRRPREGSLSQPPLOTQACPESSCLREGEKQNGDSSAGDPFPPAEVPTP 113  
 Db 1317 ---SSVESTSSQEHNLNHSKSVTPASTLRSGP-----GRWKTPAAVLPTP 1360  
 QY 114 EAEILAPQCH-----DSEASKLGAPAAAGSEEWGQQOQLGKKGHR 155  
 Db 1361 VA--VSQPIRTDLPPEPPPPPHYTSDFGISMPLPPPPANQA-APQSAVAAAEK 1417  
 QY 156 RPSKKRHWKPYKLTWE-EKKKFKDEKQSLRASRIRAEWFAKQGPVA----- 201  
 Db 1418 REHQGWKEKARLEERERKRREGERKLGMRQSLNPASFSPLATQAKPEKPTLQR 1477  
 QY 202 PYNITQFLMDHDQDEP-----DLTKGLYSKRAAKSDTSDDDPMEGGSEEDGSDGM 255  
 Db 1478 PQTIV--IRELQPOQOQRTIERRDLOYITISKRELSGSGSLSPDPW----- 1521  
 QY 256 GGGGSEFLQRDSERYERYHTESLQNMKSQELIKYLELEKCLSMEDENNRLR----- 309  
 Db 1522 ---KRDAREKLEKQOQNHIVDM---LSKEIHELQNGKDRTAESDRKLKMLEW 1569  
 QY 310 ---LESKRLGSDDARVRELEL---LDRLEAENLQLTNEHLHQOEAPLSK 356  
 Db 1570 QPQKRLQESKQDEDDDEEDDDVDVMTLMQRLBAERARLQDEERRRQOQLEBRK 1626

RESULT 11  
 US-07-814-964-11

```

Sequence 11, Application US/0781964
Patent No. 5359047
GENERAL INFORMATION:
APPLICANT: Donahue, Brian A.
APPLICANT: Toney, Jeffrey H.
APPLICANT: Bruhn, Suzanne L.
APPLICANT: Pil, Pieter M.
APPLICANT: Brown, Steven
APPLICANT: Kelleit, Patti
APPLICANT: Essigmann, John M.
APPLICANT: Lipard, Stephen J.
TITLE OF INVENTION: DNA Structure Specific Recognition
TITLE OF INVENTION: Protein and Uses Therefor
NUMBER OF SEQUENCES: 13
CORRESPONDENCE ADDRESS:
ADDRESS: Hamilton, Brook, Smith & Reynolds, P.C.
STREET: 2 Millitia Drive
CITY: Lexington
STATE: MA
COUNTRY: USA
ZIP: 02173
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patentin Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/07/814,964
FILING DATE: 19911226
CLASSIFICATION: 435
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 07/539,906
FILING DATE: 18-JUN-1990
ATTORNEY/AGENT INFORMATION:
NAME: Granahan, Patricia
REGISTRATION NUMBER: 32,227
REFERENCE/DOCKET NUMBER: MIT-4787AAA
TELECOMMUNICATION INFORMATION:
TELEPHONE: 617-861-6240
TELEFAX: 617-861-9540
INFORMATION FOR SEQ ID NO: 11
SEQUENCE CHARACTERISTICS:
LENGTH: 723 amino acids
TYPE: AMINO ACID
TOPOLOGY: linear
MOLECULE TYPE: protein
ORIGINAL SOURCE:
ORGANISM: Drosophila melanogaster
IMMEDIATE SOURCE:
CLONE: Drosophila SSRP (predicted)
FEATURE:
NAME/KEY: Domain
LOCATION: 458..507
OTHER INFORMATION: /label= Acidic
FEATURE:
NAME/KEY: Domain
LOCATION: 518..547
OTHER INFORMATION: /label= Basic I
FEATURE:
NAME/KEY: Domain
LOCATION: 547..620
OTHER INFORMATION: /label= HMG-box
FEATURE:
NAME/KEY: Domain
LOCATION: 632..649
OTHER INFORMATION: /label= Basic II
FEATURE:
NAME/KEY: Domain
LOCATION: 657..723
OTHER INFORMATION: /label= Mixed Charge
US-07-814-964-11

```

```

Query Match 6.4%; Score 122; DB 1; Length 723;
Best Local Similarity 20.4%; Pred. No. 0.031;
Matches 61; Conservative 55; Mismatches 125; Indels 58; Gaps 9;

Qy 63 GEGSLSEPPLOTQACPPSSCLREGKQNGDSSAGSDPP---PAVEPTPEELIA 119
Db 441 GDSNDENEDPAYIARAKAREKEEDDDGDSDESDTDFRPNESVVAEYDSNV 500
Qy 120 QPCDSEASKGLGAPAAAGEEWGQQOLGKKHRRRPSKKRWKPYVYKLTWBEKKPD 179
Db 501 DSDDDSDAS-----GGGSDSDAKKKKE---KKSEKKKKKKH-KEXRTKPSKKKD 551
Qy 180 EKSLASRIRIAPMEKQPVAPYNTTQFLMDHD-----QEPDLKGLYSKR----- 228
Db 552 SGKPKRA-----TTAFMLWINDRESIKRENPGIKVETIAKKGGEMW 593
Qy 229 -----AAKSDTSDDDFMEEGEGDGGSDGSGDSEFLQRFSTYERHYTE 277
Db 594 KEIKDKSKMEDAAAKQKQRYHDEMRYKPEAGSDSNEKGGKSKKRTKPSKKAATS 653
Qy 278 SLQNSKQELIKYELKCLSRMEDENNR-LRLSKRLGDDARVRELELRLRAE 335
Db 654 GSGFKS-----KEYISDDSDTSSDDEKQNEPAKKSKSPSDGAKKKAKKSESPRESE 707

```

```

RESULT 12
US-08-258-442-11
Sequence 11, Application US/08258442
Patent No. 5670621
GENERAL INFORMATION:
APPLICANT: Donahue, Brian A.
APPLICANT: Toney, Jeffrey H.
APPLICANT: Bruhn, Suzanne L.
APPLICANT: Pil, Pieter M.
APPLICANT: Brown, Steven
APPLICANT: Kelleit, Patti
APPLICANT: Essigmann, John M.
APPLICANT: Lipard, Stephen J.
TITLE OF INVENTION: DNA Structure Specific Recognition
TITLE OF INVENTION: Protein and Uses Therefor
NUMBER OF SEQUENCES: 13
CORRESPONDENCE ADDRESS:
ADDRESS: Hamilton, Brook, Smith & Reynolds, P.C.
STREET: 2 Millitia Drive
CITY: Lexington
STATE: MA
COUNTRY: USA
ZIP: 02173
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patentin Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/258,442
FILING DATE:
CLASSIFICATION: 530
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 07/539,906
FILING DATE: 18-JUN-1990
ATTORNEY/AGENT INFORMATION:
NAME: Granahan, Patricia
REGISTRATION NUMBER: 32,227
REFERENCE/DOCKET NUMBER: MIT-4787AAA
TELECOMMUNICATION INFORMATION:
TELEPHONE: 617-861-6240
TELEFAX: 617-861-9540
INFORMATION FOR SEQ ID NO: 11:
SEQUENCE CHARACTERISTICS:
LENGTH: 723 amino acids
TYPE: amino acid
TOPOLOGY: linear

```

```

? MOLECULE TYPE: protein
? ORIGINAL SOURCE:
? ORGANISM: Drosophila melanogaster
? IMMEDIATE SOURCE:
? CLONE: Drosophila SSBP (predicted)
? FEATURE:
? NAME/KEY: Domain
? LOCATION: 458..507
? OTHER INFORMATION:
? FEATURE:
? NAME/KEY: Domain
? LOCATION: 518..547
? OTHER INFORMATION: /label= Basic I
? FEATURE:
? NAME/KEY: Domain
? LOCATION: 547..620
? OTHER INFORMATION:
? FEATURE:
? NAME/KEY: Domain
? LOCATION: 632..649
? OTHER INFORMATION: /label= Basic II
? FEATURE:
? NAME/KEY: Domain
? LOCATION: 657..723
? OTHER INFORMATION: /label= Mixed Charge
? OS-08-258-442-11

```

```

Query Match 6.4%: Score 122; DB 1; Length 723;
Best Local Similarity 20.4%: Pred. No. 0.031;
Matches 61; Conservative 55; Mismatches 125; Indels 58; Gaps 9

QY 63 GEGSLSEOPPLQTOACPESSCLAREGEKQNGDDSSAGDFPP--PAVEEFTPAEALLA 119
Db GDSDNENEPDPAVYLARLKAREKEEDDDGDSDESTDDEDFKPNENESDVAAEEYSNYES 500
QY 120 QPCDSEASKGAPAAAGEEEMGGQOQRLGKKHRRSPKRRHMKPPYKLTWEKKYFD 179
Db DSDDDSDAS---GGGSDSDQAKKKK---KSKSEKKEKEKH-KEKERTKPSKKKD 551
QY 180 EKQSLRASRIAEWMPKQPVAVYTTQFPLMDHD-----QEEPDLKTGLYSK- 228
Db 552 SGRKRA-----TTAPMLMLNDTRESIKRENIGIKYTLAKKGGEW 593
QY 229 -----AAAKSDTSDDFMEEGEEDGSDGMGDSSEFLQRPSEYRYHTE 277
Db 594 KELDKSKWEADAAAKDKQRHYDEMNNYKPEAGGSDSDENEKGKSSKKRKTPEPSKKANTS 653
QY 278 SLQNNKQELIKETYLEKLCISRWEDENNR-LRLFSKRLGDDAVRLELELDLRAS 335
Db 654 GSGFRS-----KEYISDDDTSTSDDEKNEPAKKSKPSDDDAKKKAKKASESEPESE 707

RESULT 13
US-08-328-809-6
; Sequence 6, Application US/08328809
; Patent No. 5705334
; GENERAL INFORMATION:
; APPLICANT: Lippard, Stephen J.
; APPLICANT: Essigmann, John M.
; APPLICANT: Donahue, Brian A.
; APPLICANT: Toney, Jeffrey H.
; APPLICANT: Bruhn, Suzanne L.
; APPLICANT: Pil, Pieter M.
; APPLICANT: Brown, Steven
; APPLICANT: Kelleck, Patti
; TITLE OF INVENTION: Uses For DNA Structure-Specific
; NUMBER OF SEQUENCES: 8
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Patent Administrator, Testa, Hurwitz & Thibault
; STREET: 53 State Street
; CITY: Boston
; STATE: MA
;

```

```

: COUNTRY: USA
: ZIP: 02109
: COMPUTER READABLE FORM:
: MEDIUM TYPE: Floppy disk
: COMPUTER: IBM PC compatible
: OPERATING SYSTEM: PC-DOS/MS-DOS
: SOFTWARE: Patent in Release #1.0, Version #1.25
: CURRENT APPLICATION DATA:
: APPLICATION NUMBER: US/08/328,809
: FILING DATE:
: CLASSIFICATION: 435
: ATTORNEY/AGENT INFORMATION:
: NAME: Fenton, Gillian M.
: REGISTRATION NUMBER: 36,508
: REFERENCE/DOCKET NUMBER: MIT-023 (5473/24)
: TELECOMMUNICATION INFORMATION:
: TELEPHONE: 617-248-7000
: TELEFAX: 617-248-7100
: INFORMATION FOR SEQ ID NO: 6:
: SEQUENCE CHARACTERISTICS:
: LENGTH: 723 amino acids
: TYPE: amino acid
: TOPOLOGY: linear
: MOLECULE TYPE: protein
: ORIGINAL SOURCE:
: ORGANISM: Drosophila melanogaster
: IMMEDIATE SOURCE:
: CLONE: Drosophila SSRP (predicted)
: FEATURE:
: NAME/KEY: Domain
: LOCATION: 458..507
: OTHER INFORMATION: /label= Acidic
: FEATURE:
: NAME/KEY: Domain
: LOCATION: 518..547
: OTHER INFORMATION: /label= Basic I
: FEATURE:
: NAME/KEY: Domain
: LOCATION: 547..620
: OTHER INFORMATION: /label= HMG
: FEATURE:
: NAME/KEY: Domain
: LOCATION: 632..649
: OTHER INFORMATION: /label= Basic II
: FEATURE:
: NAME/KEY: Domain
: LOCATION: 657..723
: OTHER INFORMATION: /label= Mixed Charge
: US-08-328-809-6
:
: Query Match 6.4%; Score 122; DB 1; Length 723;
: Best Local Similarity 20.4%; Pred. No. 0.031;
: Matches 61; Conservative 55; Mismatches 125; Indels 58; Gaps 9;
:
: QY GEGSLBQPPPLTOTCPBESSCLREGKGNGDSDSAGDFPP--PAVEPTAEALLA 119
: Db : : : : : : : : : : : : : : : : : : : : : : : : : : : :
: 441 GDSNENNEPAYIARLVAEAREKEDEDDGSDSEESTDEDFKPNENSDVAEEYSNVS 500
: QY 120 QPCHDSASIKLGAIPAAGGEEMGOQROLGKKRRRPRSKKRWKPYVYKLTWEKKKFD 179
: Db 501 DSDDDSDAS-----GGGSDSDGAKKKK--KSEKKKEKKKH-KEKERYKKSKKKKD 551
: QY 180 EKOSLRASRIAEWFAKGQVAPYNTTQFLMDHD-----QEEPDIKTGLYSKR----- 228
: Db 552 SKRPKRA-----TTAFMLMINDTRESIKRENPGLKYTEIAKGGEMW 593
: QY 229 -----AAKSDDTSDDDFMEEGGEEDGSGDGMGDDGSEFLORDSETYERYHTE 277
: Db 594 KELDKSKWEDAAAKQQRVHDEWRNYKPPAGGSDNEKGGKSKSKRRTESPSSKANTS 653
: QY 278 SLQMSKQELIKYLYLEKLSMENNRR-LRLESKRIGGDARVRELEFELRLAE 335
: Db 654 GSGFKS-----KEYISDDSDTSSDDEKDNBPAAKKSKRPSPDGDAAKKKASSEPSSE 707

```

```

RESULT 14
US-08-866-840-6
; Sequence 6, Application US/08866840
; Patent No. 6475791
; GENERAL INFORMATION:
; APPLICANT: Lippard, Stephen J.
; APPLICANT: Essigmann, John M.
; APPLICANT: Donahue, Brian A.
; APPLICANT: Toney, Jeffrey H.
; APPLICANT: Bruhn, Suzanne L.
; APPLICANT: Pil, Pieter M.
; APPLICANT: Brown, Steven
; APPLICANT: Kellelt, Patci
; TITLE OF INVENTION: Uses For DNA Structure-Specific
; NUMBER OF SEQUENCES: 8
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Patent Administrator, Testa, Hurwitz & Thibault
; STREET: 53 State Street
; CITY: Boston
; STATE: MA
; COUNTRY: USA
; ZIP: 02109
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; OPERATING SYSTEM: IBM PC compatible
; SOFTWARE: PC-DOS/MS-DOS
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/866,840
; FILING DATE: 02-JUN-1997
; CLASSIFICATION: 435
; ATTORNEY/AGENT INFORMATION:
; NAME: Fenton, Gillian M.
; REGISTRATION NUMBER: 36,508
; REFERENCE/DOCKET NUMBER: MIT-023 (5473/24)
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 617-248-7000
; TELEFAX: 617-248-7100
; INFORMATION FOR SEQ ID NO: 6:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 723 amino acids
; TYPE: amino acid
; TOPOLOGY: linear
; MOLECULE TYPE: protein
; ORIGINAL SOURCE:
; ORGANISM: Drosophila melanogaster
; IMMEDIATE SOURCE:
; CLONE: Drosophila SSRP (predicted)
; FEATURE:
; NAME/KEY: Domain
; LOCATION: 458..507
; OTHER INFORMATION: /label= Acidic
; FEATURE:
; NAME/KEY: Domain
; LOCATION: 518..547
; OTHER INFORMATION: /label= Basic I
; FEATURE:
; NAME/KEY: Domain
; LOCATION: 547..620
; OTHER INFORMATION: /label= HMG
; FEATURE:
; NAME/KEY: Domain
; LOCATION: 632..649
; OTHER INFORMATION: /label= Basic II
; NAME/KEY: Domain
; LOCATION: 657..723
; OTHER INFORMATION: /label= Mixed Charge
US-08-866-840-6

```

```

Query Match 6.4%; Score 122; DB 4; Length 723;
Best Local Similarity 20.4%; Pred. No. 0.031;
Matches 61; Conservative 55; Mismatches 125; Indels 58; Gaps 9;

Qy 63 GEGSLSEQPPPLTOTACPPSSCLREGEKQNGDDSGADFPF---PAVEPTPEELLA 119
Db 441 GDSNENEPDPAVIAIKAEAREKEEDDDGSDSESTDEDFPENESVVAERYDSNVS 500
Qy 120 QPCHDSEASKLGAPALAGEEEMGOOQROLGKKHRRRPSKTRHWPYKLTWEEKKPD 179
Db 501 DSDDDSDAS-----GGGSDSDAKKKE---KKSEKKEKKEKH-KEXRTKPSKKKD 551
Qy 180 EKQSLPASIRAEWFAKQGVAPYNTTQFLMDHD-----QEPDLKTGLYSKR----- 228
Db 552 SGKPKRA-----TTAFMLMINDRESIKRNPGLKYTEIAKKGGEMW 593
Qy 229 -----AAKSDTSDDDFMESGEEDGSGMGDDSEFLQRFSETYERYHTE 277
Db 594 KEIKDKSKWEDAAKQKQRYHDEMRYKPEAGSDSNKGKSSKKRKTPEPSKANTF 653
Qy 278 SLQNSKQELIKLEYLEKCLSRMEDENNR-LRLSKRLGDDARVRELELDRLRAE 335
Db 654 GSGFKS-----KEYISDDSTSSDDEKNEPAKKSKPSDGDAKKKKSKSESEPESE 707

```

```

RESULT 15
PCT-US92-11107-11
; Sequence 11, Application PC/TUS9211107
; GENERAL INFORMATION:
; APPLICANT: Donahue, Brian A.
; APPLICANT: Toney, Jeffrey H.
; APPLICANT: Bruhn, Suzanne L.
; APPLICANT: Pil, Pieter M.
; APPLICANT: Brown, Steven
; APPLICANT: Kellelt, Patci
; APPLICANT: Essigmann, John M.
; TITLE OF INVENTION: DNA Structure Specific Recognition
; TITLE OF INVENTION: Protein and Uses Therefor
; NUMBER OF SEQUENCES: 13
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Hamilton, Brook, Smith & Reynolds, P.C.
; STREET: 2 Militia Drive
; CITY: Lexington
; STATE: MA
; COUNTRY: USA
; ZIP: 02173
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; OPERATING SYSTEM: IBM PC compatible
; SOFTWARE: PC-DOS/MS-DOS
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: PCT/US92/11107
; FILING DATE: 19921218
; CLASSIFICATION:
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 07/539,906
; FILING DATE: 18-JUN-1990
; ATTORNEY/AGENT INFORMATION:
; NAME: Granahan, Patricia
; REGISTRATION NUMBER: 32,227
; REFERENCE/DOCKET NUMBER: MIT-4787AAA
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 617-861-6240
; TELEFAX: 617-861-9540
; INFORMATION FOR SEQ ID NO: 11:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 723 amino acids
; TYPE: AMINO ACID
; TOPOLOGY: linear
; MOLECULE TYPE: protein
; ORIGINAL SOURCE:

```

```

; ORGANISM: Drosophila melanogaster
; IMMEDIATE SOURCE:
; CLONE: Drosophila SSRP (predicted)
; FEATURE:
; NAME/KEY: Domain
; LOCATION: 458..507
; OTHER INFORMATION: /label= Acidic
; FEATURE:
; NAME/KEY: Domain
; LOCATION: 518..547
; OTHER INFORMATION: /label= Basic I
; FEATURE:
; NAME/KEY: Domain
; LOCATION: 547..620
; OTHER INFORMATION: /label= HWG-box
; FEATURE:
; NAME/KEY: Domain
; LOCATION: 632..649
; OTHER INFORMATION: /label= Basic II
; FEATURE:
; NAME/KEY: Domain
; LOCATION: 657..723
; OTHER INFORMATION: /label= Mixed Charge
; PCT-US92-11107-11

```

Query Match 6.4%; Score 122; DB 5; Length 723;

Best Local Similarity 20.4%; Pred. No. 0.031;

```

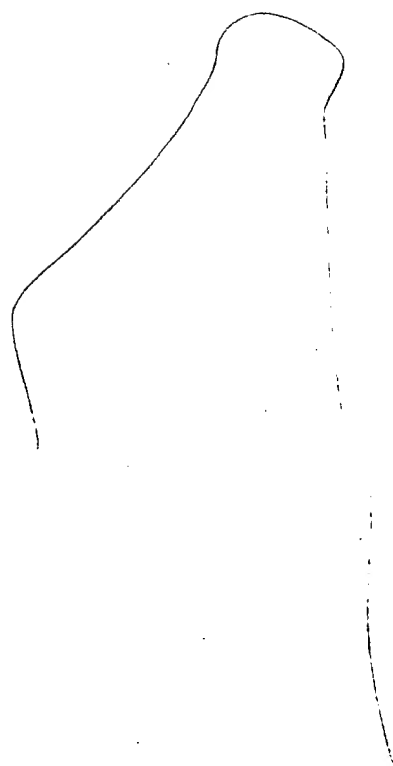
Matches 61; Conservative 55; Mismatches 125; Indels 58; Gaps 9;

QY 63 GEGSLESQPPLOTQACPESSCLREGEKQNGDDSSAGDPP--PAVEYPTPEALLA 119
Db 441 GDSDNENEPDVALRLKAEAREKEEDDDSDSESTDEDFKPNENESDVAEEYDSNVEG 500
QY 120 QPCHDSASTLGAPAGGEEMGQOQOLGKXRRRPSKKRHWKPYKLTWEKKKPD 179
Db 501 DSDDDSDAS---GGGSDSDGAKKKK---KKSEKKEKKEKH-KEKERTKKPSKKKD 551
QY 180 EKOSLRASRIAEWFAKQGVAPYNTTQFLMDHD-----QEPDLKTGLYSKR----- 228
Db 552 SGKPKRA-----TTAFMLMNDTRESIKRINPGIKVTEIAKKGEMW 593
QY 229 -----AAKSDDTSDDDFMEEGEGEDGSDGMGGSGSEFLQPDFSETRYHTE 277
Db 594 KELKDSKWEADAADKQRYHDEWRNYKPEAGGSDNEKGKSKKRKTEPSPSKANTS 653
QY 278 SLQMSKQELKEYLELEKCLSRMEDENR-LRLSEKRLGDDARVARELELELDRLRAE 335
Db 654 GSGFKS-----KEYISDDSTSSDDEKDNPAKKSKPSPDGAKKKAKSESEPESE 707

```

Search completed: February 5, 2004, 13:32:37  
 Job time : 29 secs

THIS PAGE BLANK (USPTO)



GenCore version 5.1.6  
Copyright (c) 1993 - 2004 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: February 5, 2004, 09:56:07 ; Search time 97 Seconds  
(without alignments)  
4914.369 Million cell updates/sec

Title: US-09-972-758a-1  
Perfect score: 1080  
Sequence: 1 atgcgcagcagcattctgtc.....ttcccaagtttgagagactag 1080

Scoring table: IDENTITY NUC  
Gapop 10.0, Gapext 1.0

Searched: 569978 seqs, 220691566 residues

Total number of hits satisfying chosen parameters: 1139956

Minimum DB seq length: 0  
Maximum DB seq length: 200000000

Post-processing: Minimum Match 0%  
Maximum Match 100%  
Listing first 45 summaries

Database :  
1: Issued Patents NA:\*  
2: /cgn2\_6/prodata/1/ina/5A\_COMB.seq:\*  
3: /cgn2\_6/prodata/1/ina/5B\_COMB.seq:\*  
4: /cgn2\_6/prodata/1/ina/5A\_COMB.seq:\*  
5: /cgn2\_6/prodata/1/ina/PCTUS\_COMB.seq:\*  
6: /cgn2\_6/prodata/1/ina/backfillseq.seq:\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
C 1	69	6.4	7218	1	US-08-232-463-14
C 2	68.4	6.3	1926	4	US-09-249-585A-2
C 3	68.4	6.3	1926	4	US-09-410-399-3
C 4	68.4	6.3	2580	4	US-09-050-863-2
C 5	68.4	6.3	2580	4	US-09-359-081-2
C 6	68.4	6.3	5452	2	US-09-130-114-1
C 7	68.4	6.3	8705	4	US-09-647-344A-14
C 8	68.4	6.3	9600	3	US-08-910-647-1
C 9	68.4	6.3	9600	4	US-09-620-925-1
C 10	68.4	6.3	10596	1	US-07-884-811-15
C 11	68.4	6.3	10596	1	US-07-885-971-15
C 12	68.4	6.3	10596	1	US-08-087-783A-15
C 13	68.4	6.3	10596	1	US-08-194-088B-15
C 14	68.4	6.3	10596	2	US-08-194-087-15
C 15	68.4	6.3	10596	5	PCT-US93-04648-15
C 16	67.4	6.2	1926	4	US-09-249-585A-4
C 17	67.4	6.2	1921	2	US-09-130-114-2
C 18	65.8	6.1	9531	1	US-08-056-200-93
C 19	65.8	6.1	9551	2	US-08-800-644-93
C 20	63.4	5.9	3489	2	US-08-728-323A-1
C 21	63.4	5.9	3489	4	US-09-298-568-1
C 22	63.4	5.9	3489	4	US-09-410-399-1
C 23	63.4	5.9	32207	2	US-08-757-669A-20
C 24	63.4	5.9	32207	3	US-09-230-371A-20
C 25	60.2	5.6	2793	1	US-08-209-747-1
C 26	60.2	5.6	2793	1	US-08-458-298-1
C 27	60.2	5.6	2793	1	US-08-458-298-1

28	58.4	5.4	2313	4	US-09-370-838-157	Sequence 157, App
C 29	53.4	4.9	1037	4	US-09-181-585-3	Sequence 3, Appl
C 30	53.4	4.9	1159	4	US-09-181-585-1	Sequence 1, Appl
C 31	53.4	4.9	1471	4	US-09-181-585-2	Sequence 2, Appl
C 32	51.8	4.8	1383	4	US-09-252-991A-16519	Sequence 16519, A
C 33	51.8	4.8	2781	4	US-09-252-991A-15980	Sequence 15980, A
C 34	51.8	4.8	2952	4	US-09-252-991A-16297	Sequence 16297, A
C 35	51.2	4.7	397	3	US-09-253-691-3	Sequence 3, Appl
C 36	51	4.7	1995	2	US-08-425-069-3	Sequence 3, Appl
C 37	51	4.7	1995	2	US-08-317-844B-3	Sequence 209, App
C 38	50.8	4.7	51259	3	US-08-781-891-209	Sequence 209, App
C 39	50.8	4.7	51259	4	US-09-618-156-209	Sequence 1, Appl
C 40	50.2	4.6	1949	4	US-09-740-027-1	Sequence 12655, A
C 41	49.8	4.6	1455	4	US-09-252-991A-12655	Sequence 13312, A
C 42	49.8	4.6	1488	4	US-09-252-991A-13312	Sequence 1, Appl
C 43	49.6	4.6	3138	4	US-09-434-408-1	Sequence 8476, Ap
C 44	49.4	4.6	876	4	US-09-252-991A-8476	Sequence 8333, Ap
C 45	49.4	4.6	1776	4	US-09-252-991A-8333	

ALIGNMENTS

RESULT 1  
US-08-232-463-14/c  
Sequence 14, Application US/08232463  
Patent No. 5670367  
GENERAL INFORMATION:  
APPLICANT: DORNER, F.  
APPLICANT: SCHEIFLINGER, F.  
APPLICANT: FALKNER, F. G.  
TITLE OF INVENTION: RECOMBINANT FOWLPOX VIRUS  
NUMBER OF SEQUENCES: 52  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: Foley & Lardner  
STREET: 1800 Diagonal Road, Suite 500  
CITY: Alexandria  
STATE: VA  
COUNTRY: USA  
ZIP: 22313-0299  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: Patent Release #1.0, Version #1.25  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/08/232,463  
FILING DATE:  
CLASSIFICATION: 435  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: US/07/935,313  
FILING DATE:  
APPLICATION NUMBER: EP 91 114 300.6  
ATTORNEY/AGENT INFORMATION:  
NAME: BENT, Stephen A.  
REGISTRATION NUMBER: 29, 768  
REFERENCE/DOCKET NUMBER: 30472/114 IMMU  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: (703) 836-9300  
TELEFAX: (703) 683-4109  
TELEX: 899149  
INFORMATION FOR SEQ ID NO: 14:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 7218 base pairs  
TYPE: nucleic acid  
STRANDEDNESS: single  
TOPOLOGY: linear  
IMMEDIATE SOURCE:  
CLONE: PTZgpc-F15  
US-08-232-463-14  
Query Match 6.4%; Score 69; DB 1; Length 7218;

Best Local Similarity 3.1%; Pred. No. 2.4e-06;  
Matches 12; Conservative 238; Mismatches 143; Indels 0; Gaps 0;  
QY 386 AGTTGGGGGTCCTCCGCGGCGAGAGAGAGTGGGACGACGACGACGACTGG 445  
DB 1453 AGATAGAAATTGTCACRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRR 1394  
QY 446 GGAAGAAAAATAGAGACGCCCTCAAGAGAGCGCATTTGAAACCTACTACA 505  
DB 1393 RR 1334  
QY 506 AGCTGAAGTGGAGAGAGAAAAAGTTCAAGAGAAACGAGCTTCCAGCTTCAAGA 565  
DB 1333 RR 1274  
QY 566 TCCGAGCGAGATGTTGCGACAGGCGCGGCGCTATTAACACACGAGTTCC 625  
DB 1273 RR 1214  
QY 626 TCATGATGATCAAGACGAGAGAGCGGATCTCAAAACCGGCTGTACTCCAGCGG 685  
DB 1213 RR 1154  
QY 686 CCGCGCCCAATCCGACGACACGAGATGACATTGAGAAAGGGGTGAGAGG 745  
DB 1153 RR 1094  
QY 746 ATGGGGCGAGCATGGATGGAGGAGGAGCGCA 778  
DB 1093 RR 1061

## RESULT 2

US-09-249-585A-2  
; Sequence 2, Application US/09249585A  
; Patent No. 6417002  
; GENERAL INFORMATION:  
; APPLICANT: Horlick, Robert  
; TITLE OF INVENTION: METHOD FOR MAINTENANCE AND SELECTION OF EPISOMES  
; FILE REFERENCE: 0867/0D905  
; CURRENT APPLICATION NUMBER: US/09/249,585A  
; CURRENT FILING DATE: 1999-02-11  
; NUMBER OF SEQ ID NOS: 18  
; SOFTWARE: Patent version 3.0  
; SEQ ID NO 2  
; LENGTH: 1926  
; TYPE: DNA  
; ORGANISM: Epstein Barr Virus  
; FEATURE:  
; NAME/KEY: CDS  
; LOCATION: (1)..(1926)  
; OTHER INFORMATION: coding strand of EBNA-1 DNA  
US-09-249-585A-2

Query Match 6.3%; Score 68.4; DB 4; Length 1926;  
Best Local Similarity 44.0%; Pred. No. 2.4e-06;  
Matches 288; Conservative 0; Mismatches 366; Indels 0; Gaps 0;

QY 406 GGGGGCGAAGAGAGTGGGACGACGACGACGAGTGGGGAAGAAAACTTAAGAGA 465  
DB 305 GGGCAGAGAGAGAGAGGGGCGAGAGACAGAGAGGGGCGAGAGGGGCGAG 364  
QY 466 CGCCCGTCGAAGAGAGCGGCAATTGAAACCGTACTCAAGCTGAAGTGGAGAGAG 525  
DB 365 GAGGGCAGAGAGAGAGGGGCGAGAGACAGAGAGGGGCGAGAGGGGCGAG 424  
QY 526 AAAAAGTTGACAGAGAAACAGAGCTTCAAGCTTCAAGATCCGAGCGAGATGTTCCGC 585  
DB 425 CAGAGAGAGAGAGGGGCGAGAGAGAGAGAGAGGGGCGAGAGGGGCGAGAGAG 484  
QY 586 AAGGGCCAGCGCGTCCGCTTAAACACGAGATTTCTCATGATGATCAAGACGAG 645  
DB 485 GGGCAGAGAGAGAGGGGCGAGAGAGAGAGAGGGGCGAGAGAGAGAGAGGGGCGAG 544

QY 646 GAGAGCCGAGATCTCAAAACCGGCTGTATCTCAAGCGGCGCGCCAAATCCAGCAG 705  
DB 545 GAGGGCAGAGACAGAGAGAGGGGCGAGAGGGGCGAGAGGGGCGAGAGAGAGAGGGG 604  
QY 706 ACCAGCATGATGATCTTCAAGAGAGAGGGGTAGAGAGATGGGGCGACGATGGATG 765  
DB 605 CAGAGCAGAGAGAGAGAGGGGCGAGAGGGGCGAGAGCAGAGAGGGGCGAGAGAG 664  
QY 766 GAGAGGAGCGGACGAGATTTCTGACGCGGGACTTCTCGAGACGTAACGAGCGTAC 825  
DB 665 GAGGGCAGAGAGAGGGGCGAGAGGGGCGAGAGCGAGAGGGGCGAGAGAGGGGCGAG 724  
QY 826 ACGAGAGCTTCAGAAATGACAGAGAGAGAGCTCATCAAGAGTACTTGAATGAGAG 885  
DB 725 CAGAGAGGCGAGAGAGAGAGGGGCGAGAGGGGCGAGAGCAGAGAGGGGCGAGAG 784  
QY 886 AAGTCCCTTCGCGCATGAGAGAGAGAAACCGGCTGCGCTGAGAGCAAGCGCTG 945  
DB 785 GAGCAGAGAGGGGCGAGAGGGGCGAGAGCGAGAGAGGGGCGAGAGAGAGAGAG 844  
QY 946 GGTGGCGAGAGCGCGGTGTGGGGAGCTGAGCTGAGCTGAGACCGGCTGGCGCGAG 1005  
DB 845 GAGGGCAGAGAGGGGCGAGAGAGAGGGGCGAGAGGGGCGAGAGAGAGAGAGGGGCGAG 904  
QY 1006 AACCTTCAGCTCTGACGAGAGAGAACTGCAACCGGCGAGAGAGAGAGCGCG 1059  
DB 905 GGGCAGAGACAGAGAGGGGCGAGAGGGGCGAGAGCGAGAGAGAGAGAGAGAGAGAG 958

## RESULT 3

US-09-410-399-3  
; Sequence 3, Application US/09410399  
; Patent No. 6482587  
; GENERAL INFORMATION:  
; APPLICANT: Robertson, Erle S.  
; APPLICANT: Colter, Murray A.  
; TITLE OF INVENTION: Methods to Inhibit or Enhance the Binding of Viral DNA  
; FILE REFERENCE: UM-03778  
; CURRENT APPLICATION NUMBER: US/09/410,399  
; CURRENT FILING DATE: 1999-10-01  
; NUMBER OF SEQ ID NOS: 6  
; SOFTWARE: Patent version 2.0  
; SEQ ID NO 3  
; LENGTH: 1926  
; TYPE: DNA  
; ORGANISM: Epstein-Barr Virus  
US-09-410-399-3

Query Match 6.3%; Score 68.4; DB 4; Length 1926;  
Best Local Similarity 44.0%; Pred. No. 2.4e-06;  
Matches 288; Conservative 0; Mismatches 366; Indels 0; Gaps 0;

QY 406 GGGGGCGAAGAGAGTGGGACGACGACGACGAGTGGGGAAGAAAACTTAAGAGA 465  
DB 305 GGGCAGAGAGAGAGAGGGGCGAGAGAGAGAGAGGGGCGAGAGGGGCGAGAGGGGCGAG 364  
QY 466 CGCCCGTCGAAGAGAGCGGCAATTGAAACCGTACTCAAGCTGAAGTGGAGAGAGAG 525  
DB 365 GAGGGCAGAGAGAGAGGGGCGAGAGCGAGAGAGGGGCGAGAGGGGCGAGAGGGGCGAG 424  
QY 526 AAAAAGTTGACAGAGAAACAGAGCTTCAAGCTTCAAGATCCGAGCGAGATGTTCCGC 585  
DB 425 CAGAGCAGAGAGAGGGGCGAGAGAGAGAGAGGGGCGAGAGGGGCGAGAGAGAGAGAG 484  
QY 586 AAGGGCCAGCGCGTCCGCTTAAACACGAGATTTCTCATGATGATCAAGACGAG 645  
DB 485 GGGCAGAGAGAGAGGGGCGAGAGAGAGAGAGGGGCGAGAGAGAGAGAGGGGCGAGAG 544  
QY 646 GAGAGCCGAGATCTCAAAACCGGCTGTATCTCAAGCGGGGCGCGCCAAATCCAGCAG 705  
DB 545 GAGGGCAGAGACAGAGAGGGGCGAGAGGGGCGAGAGGGGCGAGAGAGAGAGAGGGGCGAG 604



QY 706 ACCAGCATGACGACTTTCATGGAAGAAGGGGTGAGAGATGGGGCAGCGATGGGATG 765  
DB 605 CAGGACAGAGAGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 664  
QY 766 GAGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 825  
DB 665 GAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 724  
QY 826 AGGAGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 885  
DB 725 CAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 784  
QY 886 AAGTGCTCTCGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 945  
DB 785 GAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 844  
QY 946 GGTGGCAGAGCGCGGTGTGCGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 1005  
DB 845 GAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 904  
QY 1006 AACCTCAGCTGTGACCGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGG 1059  
DB 905 GGGCAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGG 958

RESULT 4  
US-09-050-863-2  
Sequence 2, Application US/09050863  
Patent No. 6114111

GENERAL INFORMATION:  
APPLICANT: Lao, Yang  
APPLICANT: Hiang, Betty  
APPLICANT: Payan, Don  
TITLE OF INVENTION: Mammalian Protein Interaction Cloning  
NUMBER OF SEQUENCES: 5  
CORRESPONDENCE ADDRESS:  
ADDRESS: Flehr, Hohbach, Test, Albritton & Herbert  
STREET: 4 Embarcadero Center, Suite 3400  
CITY: San Francisco  
STATE: CA  
COUNTRY: USA  
ZIP: 94111-4187  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: Patentin Release #1.0, Version #1.30  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/09/050, 863  
FILING DATE: 30-MAR-1998  
CLASSIFICATION:  
ATTORNEY/AGENT INFORMATION:  
NAME: Silva, Robin M.  
REGISTRATION NUMBER: 38,304  
TELEPHONE: (415) 781-1989  
TELEFAX: (415) 949-8711  
INFORMATION FOR SEQ ID NO: 2:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 2580 base pairs  
TYPE: nucleic acid  
STRANDEDNESS: unknown  
TOPOLOGY: unknown  
MOLECULE TYPE: DNA  
US-09-050-863-2

Query Match 6.3%; Score 68.4; DB 3; Length 2580;  
Best Local Similarity 44.0%; Pctd. No. 2.6e-06;  
Matches 288; Conservative 0; Mismatches 366; Indels 0; Gaps 0;

QY 406 GGGGCGAAGAGAGTGGGGACAGACAGACAGCTGGGGAGAGAAAAACATAGAGA 465  
DB 688 GGGCAGAGAGCAGAGAGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 747  
QY 466 CGCCCGTCCAGAGAGAGCGGATTTGGAAACCGTACTCAAGCTGAGTGAAGAGAG 525  
DB 748 GAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 807  
QY 526 AAAAGTTGACAGAGAGAGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 585  
DB 808 CAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 867  
QY 586 AAGGCGCAGCGGTCGCGCTTATTAACACAGCAGTTCCTCATGATGATCAACAG 645  
DB 868 GGGCAGAGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 927  
QY 646 GAGGAGCGGAGTCTAAAAACCGCTGTATCTCAAGCCGGCCCGCCAAATCCGACAG 705  
DB 928 GAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 987  
QY 706 ACCAGCATGACGACTTTCATGGAAGAAGGGGTGAGAGATGGGGCAGCGATGGGATG 765  
DB 988 CAGGAGCAGAGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 1047  
QY 1048 GAGGGGACAGAGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 1107  
DB 826 ACAGAGAGCTTCAAAATGAGCAAGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 885  
QY 1108 CAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 1167  
DB 886 AAGTGCTCTCGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 945  
QY 1168 GAGCAGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 1227  
DB 946 GGTGGCAGAGCGCGGTGTGCGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 1005  
QY 1228 GAGGGGACAGAGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 1287  
DB 1006 AACCTCAGCTGTGACCGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGG 1059  
QY 1288 GGGCAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGG 1341

RESULT 5  
US-09-359-081-2  
Sequence 2, Application US/09359081  
Patent No. 6316223

GENERAL INFORMATION:  
APPLICANT: Lao, Yang  
APPLICANT: Hiang, Betty  
APPLICANT: Payan, Don  
TITLE OF INVENTION: Mammalian Protein Interaction Cloning  
NUMBER OF SEQUENCES: 5  
CORRESPONDENCE ADDRESS:  
ADDRESS: Flehr, Hohbach, Test, Albritton & Herbert  
STREET: 4 Embarcadero Center, Suite 3400  
CITY: San Francisco  
STATE: CA  
COUNTRY: USA  
ZIP: 94111-4187  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: Patentin Release #1.0, Version #1.30  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/09/359, 081  
FILING DATE: 22-JUL-1999  
CLASSIFICATION: <Unknown>  
PRIOR APPLICATION DATA:

APPLICATION NUMBER: 09/050,863  
 FILING DATE: <unknown>  
 ATTORNEY/AGENT INFORMATION:  
 NAME: Silva, Robin M.  
 REGISTRATION NUMBER: 38,304  
 REFERENCE/DOCKET NUMBER: A-65638/DJB/RMS  
 TELECOMMUNICATION INFORMATION:  
 TELEPHONE: (415) 781-1989  
 TELEFAX: (415) 949-8711  
 INFORMATION FOR SEQ ID NO: 2:  
 SEQUENCE CHARACTERISTICS:  
 LENGTH: 2580 base pairs  
 type: nucleic acid  
 STRANDEDNESS: unknown  
 TOPOLOGY: unknown  
 MOLECULE TYPE: DNA  
 SEQUENCE DESCRIPTION: SEQ ID NO: 2:  
 GS-09-359-081-2

Query Match	6.3%	Score 68.4;	DB 4;	Length 2580;
Best Local Similarity	44.0%;	Pred. No. 2.6e-06;		
Matches 288;	Conservative	0;	Mismatches 366;	Indels 0;
				Gaps 0;

QY	406	GGGGGCGAAAGAGAGTGGGGGACGACGACGAGACAGCTGGGGGAAAGAAAAACATAAGAGA	465
Db	688	GGGCGAGGAGCAGAGAGAGGGGCGAGGACGAGAGAGGGGCGAGAGGGGCGAGAGGGGCGAG	747
QY	466	CGCCGCTCCAAAGAAAGAACCGCATTTGGAAAACCGTACTCAAGCTGAACTGGGAGAGAG	525
Db	748	GAGGGGCGAGGAGCAGAGAGAGGGGGCAGGAGCGAGAGAGGGGGCAGAGAGGGGCGAGAGGGG	807
QY	526	AAAAAGTTCCAGCAGAGAAACAGAGCCTTGGAGCTTCAAGATCCGAGCCGAGATGTTCCGC	585
Db	808	CAGGAGCAGGAGAGAGGGGCGAGAGGACGAGAGAGAGGGGGCAGAGAGGGGCGAGAGGAGAG	867
QY	586	AAGGGCCAGCCGGGTGCGCCCTATTAACCAACGCGAGTTCTTCATGATGATCAACAGCAG	645
Db	868	GGCGCAGAGAGGGGCGAGAGGGGCGAGAGCGAGAGAGAGGGGGCAGAGACAGAGAGAGGGGCGAG	927
QY	646	GAGAGGCCGATTCCTCAAAAACGGCCTGTATCTTCAAGCGGGCCGCCCAATTCGAGCAC	705
Db	928	GAGGGGCGAGGAGCAGAGAGAGGGGCGAGAGGGGCGAGAGGGGCGAGAGCAGAGAGAGGGG	987
QY	706	ACCAAGCGATGACGACTTCATGAGAGAGAGGGGTGAGAGATGGGGGCGAGCGATGGGATG	765
Db	988	CAGGAGCAGAGAGGGGCGAGAGGGGCGAGAGGGGCGAGAGCGAGAGAGGGGCGAGAGGAG	1047
QY	766	GGAGGGGACGGCAGCGAGTTTCTGCAAGCGGGACTTCTCGGAGACCTGACGAGCGGTATCAC	825
Db	1048	GAGGGGCGAGAGGGGCGAGAGGGGCGAGAGCGAGAGGGGGCGAGAGCAGAGAGGGGCGAGAG	1107
QY	826	ACGGAGAGCCTTCGACAGACATAGACAGACAGAGACTCATCAAGAGTAACCTTGGAACTGGAG	885
Db	1108	CAGGAGGGGGCGAGAGCAGAGAGGGGCGAGAGGGGCGAGAGCGAGAGAGGGGCGAGAGGGGCGAG	1167
QY	886	AAGTGCTCTTCGCGCATGATGAGGACAGAAACAACCGGCTGCGGCTGGAAGCAAGGCGCTG	945
Db	1168	GAGCAGAGAGGGGCGAGAGGGGCGAGAGCGACGAGAGAGGGGGCGAGAGGGGCGAGAGCGAGAG	1222
QY	946	GGTGGCGACGACGCGCGTGTGCGGAGCTGGAAGCTGGAAGCTTGGACCGGCTGCGCGCCGAG	1003
Db	1228	GAGGGGCGAGAGGGGCGAGAGCAGGAGGGGCGAGAGGGGCGAGAGCGAGAGCAGAGGGGCGAGAG	1287
QY	1006	AACTCTCCAGCTGCTGACCGAGAACGAACTTGACACCGGCGAGCGAGAGCCGACCG	1059
Db	1288	GGGCGAGAGCAGAGAGGGGCGAGAGGGGCGAGAGCGAGAGCGAGAGAGGGGCGAGAGCGAG	1341

RESULT 6  
US-09-130-114-1/c  
; Sequence 1, Application US/09130114  
; Patent No. 5976807  
; GENERAL INFORMATION:

```

1  APPLICANT: Horlick, Robert A.
2  APPLICANT: Damaji, Baasam B.
3  APPLICANT: Robbins, Alan K.
4  TITLE OF INVENTION: Eukaryotic Cells Stably Expressing Genes
5  TITLE OF INVENTION: From Multiple Transfected Episomes
6  FILE REFERENCE: 0867/11D903US1
7  CURRENT APPLICATION NUMBER: US/09/130,114
8  CURRENT FILING DATE: 1998-08-06
9  NUMBER OF SEQ. ID NOS: 36
10 SOFTWARE: FastSeq for Windows Version 3.0
11 SEQ ID NO 1
12 LENGTH: 5452
13 TYPE: DNA
14 ORGANISM: YEENBA
15
16 US-09-130-114-1

```

Query Match	6.3%;	Score 68.4;	DB 2;	length 5452;
Best Local Similarity	44.0%;	Pred. No. 3.1e-06;		
Matches 288;	Conservative 0;	Mismatches 366;	Indels 0;	Gaps 0;

QY	406	GGGGCCAAAGAGAGTGGGGACACACACAGAGACAGCTGGGGAAAGAAAAACTTAAGAGA	465
Db	2117	GGGACAGAGCAGAGAGAGGGGCGAGAGCAGAGAGAGGGGCGAGAGGGGCGAGAGGGCGAG	2055
QY	466	CGCCCGTCCAGAAAGACGGCATTGGAAACCGTACTCAAGCTGAATGGGAAAGAGAG	525
Db	2057	GAGGGGCAGAGGCAGGAGAGGGGCGAGAGCAGAGAGAGGGGCGAGAGGGCGAGAGGGG	1998
QY	526	AAAAAGTTCCACCGAAGAAACAGAGGCTTCAGAGCTTCAGAGATCCGAGCCGAATGTTGGCC	585
Db	1997	CAGAGACAGAGAGAGGGGCGAGAGCAGAGAGAGGGGCGAGAGGGGCGAGAGACAGAGAGAG	1938
QY	586	AAGGCCACAGCCGAGTGGCGCCCTTATTAACACACGAGTTCCTCATGAGTATCAGACACAG	645
Db	1937	GGGACAGAGAGGGGCGAGAGGGGCGAGAGCAGAGAGGAGGAGGGGCGAGAGAGAGGGCGAG	1878
QY	646	GAGGAGCCGGATCTCAAAAACGGGCTGTACTCCAAGGGGCGCCGCCAAATCCGAGCAG	705
Db	1877	GAGGGGCGAGGCGCAGGAGAGGGGCGAGAGGGGCGAGAGGGGCGAGAGCAGAGAGAGGGG	1818
QY	706	ACCAAGCATGACGACCTTCATGAAAGAAAGGGGGTAGAGAGGATGGGGGACGAGTGGGATG	765
Db	1817	CAGAGACAGAGAGGGGCGAGAGGGGCGAGAGGGGCGAGAGCCAGAGAGGGGCGAGAGCAGAG	1758
QY	766	GAGGGGACGGCAGCAGGATTCTGACAGGGGACTTCCGAGAGCTACGAGCGGTACCAAC	825
Db	1757	GAGGGGCGAGAGGGGCGAGAGGGGCGAGAGCGAGAGGGGCGAGAGCGAGAGGGGCGAGAG	1638
QY	826	ACGGAAGAGCTTCGAGAAACATGAGCAACAGAGAGCTCATCAAGAGTACCTGGAACCTGGAG	885
Db	1697	CAGAGAGGGGCGAGAGCAGAGGGGCGAGAGGGGCGAGAGCCAGAGAGGGGCGAGAGAGGGGCG	1638
QY	886	AAGTGCCTTCGCGCATATGAGAGAGAGAAACAAACGGGCTGCGGCTGAGAGCAAGCGGCTG	945
Db	1637	GAGAGAGAGGGGCGAGAGGGGCGAGAGCAGAGAGAGGGGCGAGAGGGGCGAGAGCAGAGAG	1578
QY	946	GGTGGCCACGACGCGCGTGTGCGGGAGCTGAGCTGGAAGCTGGAACCGGCTGCGCGCCGAG	1005
Db	1577	GAGGGGCGAGAGGGGCGAGAGCAGAGAGGGGCGAGAGGGGCGAGAGCGAGAGGGGCGAGAG	1518
QY	1006	AACCTCCAGTGCCTGACCGAGAAACGAACCTGACCGGAGCGAGGAGCGAGCGCCG	1059
Db	1517	GGGACAGAGCAGAGAGGGGCGAGAGGGGCGAGAGCGAGAGGAGAGGGGCGAGAGCAG	1464

RESULT 7  
US-09-647-344A-14/C  
; Sequence 14, Application US/09647344A/C  
; Patent No. 6586180  
; GENERAL INFORMATION:  
; APPLICANT: Ruffner, Duane E.  
; APPLICANT: Pierce, Michael L.  
; APPLICANT: Chen, Zhidong







```

; TITLE OF INVENTION: HEPATOCYTE GROWTH FACTOR VARIANTS
; NUMBER OF SEQUENCES: 22
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Genentech, Inc.
; STREET: 460 Point San Bruno Blvd
; CITY: South San Francisco
; STATE: California
; COUNTRY: USA
; ZIP: 94080
; COMPUTER READABLE FORM:
; MEDIUM TYPE: 3.5 inch, 1.44 Mb floppy disk
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: WinPatIn (Genentech)
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/087,783A
; FILING DATE: 13-Jul-1993
; CLASSIFICATION: 435
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: 07/884811
; FILING DATE: 18-MAY-92
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: 07/885971
; FILING DATE: 18-MAY-92
; ATTORNEY/AGENT INFORMATION:
; NAME: Marschang, Diane L.
; REGISTRATION NUMBER: 35,600
; REFERENCE/DOCKET NUMBER: P0755779P1
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 415/225-5416
; TELEFAX: 415/952-9881
; TELEX: 910/371-7168
; INFORMATION FOR SEQ ID NO: 15:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 10596 base pairs
; TYPE: Nucleic Acid
; STRANDEDNESS: Single
; TOPOLOGY: Linear
; US-08-087-783A-15

Query Match      6.3%; Score 68.4; DB 1; Length 10596;
Best Local Similarity 44.0%; Pred. No. 3.6e-06;
Matches 288; Conservative 0; Mismatches 366; Indels 0; Gaps 0;

QY 406 GGGGGCGAAGAGAGTGGGGACAGCAGACAGCTGGGGAAGAAAACTAAGAG 465
DB 2268 GGGCAGAGAGAGAGAGGGGCGAGACAGAGAGAGGGCGAGAGGGGCGAG 2327
QY 466 CGCCCGTCCAGAAAGCCGCACTTGAAACCTACTCAAGCTGAACCTGGAAAGAA 525
DB 2328 GAGGGGCGAGAGAGAGAGGGGCGAGACAGAGAGAGGGGCGAGAGGGG 2387
QY 526 AAAAAGTTGACACAGAAACAGACCTTCGAGCTTCAAGATCCGAGCCGAGATGTTGCC 585
DB 2388 CAGGACAGAGAGAGGGGCGAGACAGAGAGAGAGGGGCGAGAGGGGCGAGAG 2447
QY 586 AAGGGCCAGCGGCTCGCGCTTAACACACGCACTTCTCATGTGATGATCAGACAG 645
DB 2448 GGGCAGAGAGGGGCGAGAGGGGCGAGACAGAGAGAGGGGCGAGAGGGGCGAG 2507
QY 646 GAGGAGCGGAGTTCAAAACCGGCTTGACTTCAAGCGGGCGCGCAATCCAGACAG 705
DB 2508 GAGGGGCGAGAGAGAGAGGGGCGAGAGAGGGGCGAGAGGGGCGAGAGGGG 2567
QY 706 ACCAGGATGACACTTCATGGAAGAGGGGTGAGAGAGATGGGGGCGACGATGGGATG 765
DB 2568 CAGGAGCAGAGAGGGGCGAGAGGGGCGAGAGGGGCGAGAGGGGCGAGAGGGG 2627
QY 766 GAGAGGAGCGAGCAGAGATTTCTGAGCGGAGACTTCTCGAGAGAGTACAGAGCGGTAC 825
DB 2628 GAGGGGCGAGAGGGGCGAGAGGGGCGAGAGAGAGGGGCGAGAGGGGCGAGAG 2687
QY 826 ACAGAGAGCTTGAGAAACATGACAAAGCAGAGAGCTCATCAAGAGTAACTGGAAG 885

```

```

DB 2688 CAGAGGGGCGAGAGAGAGGGGCGAGAGGGGCGAGACAGAGAGGGGCGAGAG 2747
QY 886 AAGTGCTCTCGCGATGAGAGCAGAAACACCGCTGCGCTGAGAGCAAGCGGCTG 945
DB 2748 GAGCAGAGAGGGGCGAGAGGGGCGAGAGCGAGAGAGGGGCGAGAGGGGCGAGAG 2807
QY 946 GGTGGGAGCAGACCGCGTGTGCGGAGCTGAGCTGAGACTGAGCCGCTGCGCGCGAG 1005
DB 2808 GAGGGGCGAGAGGGGCGAGAGGGGCGAGAGGGGCGAGAGGGGCGAGAGGGGCGAG 2867
QY 1006 AACCTCAGCTCTGATCCGAGAAACGAACTGCACCGGCGACAGAGAGCGCCG 1059
DB 2868 GGGCAGAGACAGAGAGGGGCGAGAGGGGCGAGAGGGGCGAGAGAGAGGGGCGAGAG 2921

RESULT 13
US-08-194-088B-15
; Sequence 15, Application US/08194088B
; Patent No. 5580963
; GENERAL INFORMATION:
; APPLICANT: Godowski, Paul J., Lokker, Nathalie A., Mark, Melanie R.
; TITLE OF INVENTION: SINGLE-CHAIN HEPATOCYTE GROWTH FACTOR VARIANTS
; CORRESPONDENCE ADDRESS:
; NUMBER OF SEQUENCES: 21
; ADDRESSEE: Genentech, Inc.
; STREET: 460 Point San Bruno Blvd
; CITY: South San Francisco
; STATE: California
; COUNTRY: USA
; ZIP: 94080
; COMPUTER READABLE FORM:
; MEDIUM TYPE: 5.25 inch, 360 Kb floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatIn (Genentech)
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/194,088B
; FILING DATE: 09-FEB-1994
; CLASSIFICATION: 530
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: 07/884811
; FILING DATE: 18-MAY-1992
; ATTORNEY/AGENT INFORMATION:
; NAME: Gallegos, R. Thomas
; REGISTRATION NUMBER: 32,692
; REFERENCE/DOCKET NUMBER: 755D1
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 415/225-2614
; TELEFAX: 415/952-9881
; TELEX: 910/371-7168
; INFORMATION FOR SEQ ID NO: 15:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 10596 bases
; TYPE: nucleic acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; US-08-194-088B-15

Query Match      6.3%; Score 68.4; DB 1; Length 10596;
Best Local Similarity 44.0%; Pred. No. 3.6e-06;
Matches 288; Conservative 0; Mismatches 366; Indels 0; Gaps 0;

QY 406 GGGGGCGAAGAGAGTGGGGACAGCAGACAGCTGGGGAAGAAAACTAAGAG 465
DB 2268 GGGCAGAGAGAGAGAGGGGCGAGACAGAGAGAGGGGCGAGAGGGGCGAGAGGGGCGAG 2327
QY 466 CGCCCGTCCAGAAAGCCGCACTTGAAACCTACTCAAGCTGAACCTGGAAAGAA 525
DB 2328 GAGGGGCGAGAGAGAGGGGCGAGAGGGGCGAGAGGGGCGAGAGGGGCGAGAGGGG 2387
QY 526 AAAAAGTTGACAGAAACAGACCTTCGAGCTTCAAGATCCGAGCCGAGATGTTGCC 585

```



CURRENT APPLICATION DATA:  
APPLICATION NUMBER: PCT/US93/04648  
FILING DATE: 19930517  
CLASSIFICATION:  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: 07/884811  
FILING DATE: 18-MAY-92  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: 07/885971  
FILING DATE: 18-MAY-92  
ATTORNEY/AGENT INFORMATION:  
NAME: Dreger, Ginger R.  
REGISTRATION NUMBER: 33,055  
REFERENCE/DOCKET NUMBER: 755, 779P1  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: 415/225-3216  
TELEFAX: 415/952-9881  
TELEX: 910/371-7168  
INFORMATION FOR SEQ ID NO: 15:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 10596 bases  
TYPE: NUCLEIC ACID  
STRANDEDNESS: single  
TOPOLOGY: linear  
PCT-US93-04648-15

Query Match 6.3%; Score 68.4; DB 5; Length 10596;

Best Local Similarity 44.0%; Pred. No. 3.6e-06;

Matches 288; Conservative 0; Mismatches 366; Indels 0; Gaps 0;

QY 406 GGGGGCGAAGAGAGTGGGACACGACGACAGAGCTGGGGAGAAAACTTAAGAGA 465  
DB 2268 GGGCAGAGAGCAGAGAGGGGCGAGAGCAGAGAGGGGCGAGAGGGGCGAGAGGGGCGAG 2327  
QY 466 CGCCCGTCCAAGAAAGCGGCACTTGGAAACCGTAACAAGCTGAAGTGGGAAGAG 525  
DB 2328 GAGGGCAGAGCAGAGAGAGGGGCGAGAGCAGAGAGAGGGGCGAGAGGGGCGAGAGGGG 2387  
QY 526 AAAAGTTCCAGCAGAAACAGAGCTTCGAGCTTCAAGATCCAGCCGAGATGTTGCC 585  
DB 2388 CAGAGCAGAGAGAGGGGCGAGAGCAGAGAGAGGGGCGAGAGGGGCGAGAGGGGCGAGAG 2447  
QY 586 AAGGGCAGCGGCTCGGCGCTTAAACCAAGCACTTCTCATGATGATCAAGACAG 645  
DB 2448 GGGCAGAGAGGGGCGAGAGGGGCGAGAGCAGAGAGGGGCGAGAGCAGAGAGGGGCGAG 2507  
QY 646 GAGAGCCGGATCTCAAAACCGGCTGTACTCAAGGGGCGCGCCCAATCCGAGAC 705  
DB 2508 GAGGGGAGAGCAGAGAGAGGGGCGAGAGGGGCGAGAGGGGCGAGAGGGGCGAGAGGGG 2567  
QY 706 ACCAGCAGATGACGACTTCATGAGAAAGGGGTGAGAGAGATGGGGCAGCGATGGGATG 765  
DB 2568 CAGAGCAGAGAGGGGCGAGAGGGGCGAGAGGGGCGAGAGGGGCGAGAGGGGCGAGAGGG 2627  
QY 766 GAGGGGAGCGGACGAGATTCTTCAGCGGGACTTCTCGAGACGTACAGCGGTACAC 825  
DB 2628 GAGGGGAGAGGGGCGAGAGGGGCGAGAGGGGCGAGAGGGGCGAGAGGGGCGAGAGGG 2687  
QY 826 ACCGAGAGCCTGAGAAACATGACCAAGCAGAGACTCATCAAGAGTACCTGGAAGTGGAG 885  
DB 2688 CAGAGAGGGGAGAGAGGAGGGGCGAGAGGGGCGAGAGGGGCGAGAGGGGCGAGAGGG 2747  
QY 886 AAGTGCTCTCGGCACTGAGAGCAGAAACACCGGCTGCGGCTGAGAGCAAGCGGCTG 945  
DB 2748 GAGCAGAGAGGGGCGAGAGGGGCGAGAGGGGCGAGAGGGGCGAGAGGGGCGAGAGGG 2807  
QY 946 GGTGGGAGCAGCGCGGTGTGCGGAGCTGGAAGCTGGAAGCTGGAAGCTGCGCGCGAG 1005  
DB 2808 GAGGGGAGAGGGGCGAGAGGGGCGAGAGGGGCGAGAGGGGCGAGAGGGGCGAGAGGG 2867  
QY 1006 AACCTCAGCTGTGACCGAGAAAGAACTGCAACCGCAGAGAGAGAGAGAGCGCCG 1059  
DB 2868 GGGCAGAGAGAGAGGGGCGAGAGGGGCGAGAGGGGCGAGAGGGGCGAGAGGGGCGAGAG 2921

Search completed: February 5, 2004, 12:05:27  
Job time : 99 secs



GenCore version 5.1.6  
Copyright (c) 1993 - 2004 Comphen Ltd.

OM nucleic - nucleic search, using sw model

Run on: February 5, 2004, 11:18:43 ; Search time 486 Seconds  
(without alignments)  
8185.851 Million cell updates/sec

Title: US-09-972-758A-1

Perfect score: 1080  
Sequence: 1 atggcgcagccattctgtc.....ttccacgttgagactag 1080

Scoring table: IDENTITY NUC  
Gapop 10.0, Gapext 1.0

Searched: 2449703 seqs, 1841816367 residues

Total number of hits satisfying chosen parameters: 4899406

Minimum DB seq length: 0  
Maximum DB seq length: 200000000

Post-processing: Minimum Match 0%  
Maximum Match 100%  
Listing first 45 summaries

Database :

- 1: /cgn2\_6/ptodata/2/pubpna/US07\_PUBCOMB.seq.\*
- 2: /cgn2\_6/ptodata/2/pubpna/PCT\_NEW\_PUB.seq.\*
- 3: /cgn2\_6/ptodata/2/pubpna/US06\_NEW\_PUB.seq.\*
- 4: /cgn2\_6/ptodata/2/pubpna/US06\_PUBCOMB.seq.\*
- 5: /cgn2\_6/ptodata/2/pubpna/US07\_NEW\_PUB.seq.\*
- 6: /cgn2\_6/ptodata/2/pubpna/PCTUS\_PUBCOMB.seq.\*
- 7: /cgn2\_6/ptodata/2/pubpna/US08\_NEW\_PUB.seq.\*
- 8: /cgn2\_6/ptodata/2/pubpna/US08\_PUBCOMB.seq.\*
- 9: /cgn2\_6/ptodata/2/pubpna/US09\_PUBCOMB.seq.\*
- 10: /cgn2\_6/ptodata/2/pubpna/US09B\_PUBCOMB.seq.\*
- 11: /cgn2\_6/ptodata/2/pubpna/US09C\_PUBCOMB.seq.\*
- 12: /cgn2\_6/ptodata/2/pubpna/US09\_NEW\_PUB.seq.\*
- 13: /cgn2\_6/ptodata/2/pubpna/US10A\_PUBCOMB.seq.\*
- 14: /cgn2\_6/ptodata/2/pubpna/US10B\_PUBCOMB.seq.\*
- 15: /cgn2\_6/ptodata/2/pubpna/US10C\_PUBCOMB.seq.\*
- 16: /cgn2\_6/ptodata/2/pubpna/US10\_NEW\_PUB.seq.\*
- 17: /cgn2\_6/ptodata/2/pubpna/US60\_NEW\_PUB.seq.\*
- 18: /cgn2\_6/ptodata/2/pubpna/US60\_PUBCOMB.seq.\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	1080	100.0	1080	US-09-972-758-1	Sequence 1, Appli
2	1076.8	99.7	2199	US-09-745-763-33	Sequence 33, Appl
3	1076.8	99.7	3624	US-10-205-823-171	Sequence 171, App
4	434.2	40.2	1083	US-10-264-237-699	Sequence 629, App
5	368.8	34.1	495	US-09-918-995-403	Sequence 403, App
6	362.8	33.6	414	US-09-983-965-4895	Sequence 4895, App
7	355.2	32.9	461	US-09-954-531-813	Sequence 813, App
8	174.8	16.2	1330	US-10-094-749-1223	Sequence 1223, App
9	173.2	16.0	1926	US-10-108-2604-1496	Sequence 1496, App
10	173.2	16.0	1926	US-10-294-804-3	Sequence 3, Appli
11	68.4	6.3	8705	US-10-291-230-14	Sequence 14, Appli
12	68.4	6.3	8705	US-10-291-249-14	Sequence 14, Appli
13	68.4	6.3	9600	US-10-278-751-1	Sequence 1, Appli
14	68.4	6.3	10233	US-10-050-898-283	Sequence 283, App
15	68.4	6.3	10285	US-10-050-902-283	Sequence 283, App

16	67.8	6.3	1548	13	US-10-322-774-5	Sequence 5, Appli
17	67.8	6.3	3194	10	US-09-814-353-20004	Sequence 20004, A
18	67.8	6.3	4270	10	US-09-880-107-2427	Sequence 2427, A
19	67.8	6.3	4270	10	US-09-954-531-1352	Sequence 1352, Ap
20	67.8	6.3	4653	13	US-10-269-909-45	Sequence 45, Appl
21	63.4	5.9	3489	13	US-10-294-804-1	Sequence 1, Appli
22	60.8	5.6	2108	10	US-09-962-832-225	Sequence 225, App
23	58.4	5.4	2313	10	US-09-738-973-157	Sequence 157, App
24	58.4	5.4	2313	10	US-09-854-133-157	Sequence 157, App
25	58.4	5.4	2313	15	US-10-144-6494-157	Sequence 157, App
26	58.4	5.4	2314	15	US-09-764-868-12	Sequence 12, Appl
27	58.4	5.4	2446	12	US-10-057-475B-10970	Sequence 10970, A
28	58.4	5.4	2446	12	US-10-154-884B-10970	Sequence 10970, A
29	58.4	5.4	2802	13	US-10-233-045-15	Sequence 35, Appl
30	53.6	5.0	607	15	US-10-043-487-122	Sequence 122, App
31	53.6	5.0	1609	15	US-10-043-487-105	Sequence 105, App
32	53.6	5.0	3705	13	US-10-116-275-115	Sequence 315, App
33	53.4	4.9	1037	12	US-10-373-667-3	Sequence 3, Appli
34	53.4	4.9	1159	12	US-10-373-667-1	Sequence 1, Appli
35	53.4	4.9	1471	12	US-10-373-667-2	Sequence 2, Appli
36	53.2	4.9	628	13	US-10-029-386-22859	Sequence 22859, A
37	52.2	4.8	3243	15	US-10-156-761-2685	Sequence 2685, Ap
38	52.2	4.8	9025608	15	US-10-156-761-1	Sequence 1, Appli
39	52	4.8	223883	15	US-10-175-523-57	Sequence 57, Appli
40	51.8	4.8	2529	12	US-10-369-453-11518	Sequence 31518, A
41	51.4	4.8	2479	12	US-10-108-260A-1966	Sequence 1966, Ap
42	51.4	4.8	1223197	13	US-10-027-632-179264	Sequence 179264, A
43	51.4	4.8	1223197	14	US-10-027-632-179264	Sequence 179264, A
44	51.2	4.7	1799	12	US-10-104-260A-1885	Sequence 1885, Ap
45	50.8	4.7	2183	12	US-10-104-047-1064	Sequence 1064, Ap

ALIGNMENTS

RESULT 1  
US-09-972-758-1  
Sequence 1, Application US/09972758  
Patent No. US20020160497A1  
GENERAL INFORMATION:  
APPLICANT: Case Western Reserve University  
APPLICANT: Montano, Monica  
APPLICANT: Wiltman, Bryan  
TITLE OF INVENTION: Suppressors of Human Breast Cancer Cell Growth  
FILE REFERENCE: 27708/04004  
CURRENT APPLICATION NUMBER: US/09/972,758  
CURRENT FILING DATE: 2001-10-05  
PRIOR APPLICATION NUMBER: US 60/238,187  
PRIOR FILING DATE: 2000-10-05  
NUMBER OF SEQ ID NOS: 2  
SOFTWARE: PatentIn version 3.0  
SEQ ID NO 1  
LENGTH: 1080  
TYPE: DNA  
ORGANISM: Homo sapiens  
US-09-972-758-1

Query Match	100.0%	Score 1080;	DB 10;	Length 1080;
Best Local Similarity	100.0%	Pred. No. 4.7e-265;		
Matches 1080;	Conservative 0;	Mismatches 0;	Indels 0;	Gaps 0;
QY	1	ATGGCGGAGCATTCTTGTGACAAATATCAACCCAGCCCTCAACTAGCACTGTACAGGT	60	
DB	1	ATGGCGGAGCATTCTTGTGACAAATATCAACCCAGCCCTCAACTAGCACTGTACAGGT	60	
QY	61	GCTGCTGCTGTCAGAGAGCTGAACCTGAGCCGCCGCCAGGCGCGGAGAGCGGGTG	120	
DB	61	GCTGCTGCTGTCAGAGAGCTGAACCTGAGCCGCCGCCAGGCGCGGAGAGCGGGTG	120	
QY	121	CCCGAGAGAGACAGTGTGAGCAATCGAGAGCGTTCCCACTTGGTGGCTGCGGGG	180	
DB	121	CCCGAGAGAGACAGTGTGAGCAATCGAGAGCGTTCCCACTTGGTGGCTGCGGGG	180	



Db 1187 AAGCGCATTTGGAAACCGTACTACAAAGCTGAGCTGGAGAAAGAAAAGTTTCAGCGAG 1246  
 Qy 541 AAACAGAGCCTTCGAGCTTCAAGATTCGAGCCGAGATGTTGCCAAAGGCGCCGCGTTC 600  
 Db 1247 AAACAGAGCCTTCGAGCTTCAAGATTCGAGCCGAGATGTTGCCAAAGGCGCCGCGTTC 1306  
 Qy 601 GGGCCCTATACACCAAGCAGTTCCTCATGATGATCAAGACCAAGAGAGCCGATCTC 660  
 Db 1307 GGGCCCTATACACCAAGCAGTTCCTCATGATGATCAAGACCAAGAGAGCCGATCTC 1366  
 Qy 661 AAAACCGGCTTCTACTCCCAAGCGCGCGCGCCCAATCCGACGACACCGAGCATGACGAC 720  
 Db 1367 AAAACCGGCTTCTACTCCCAAGCGCGCGCGCCCAATCCGACGACACCGAGCATGACGAC 1426  
 Qy 721 TTCATGAAAGAAAGGGGTGAGAGATGGGGGAGCATGGGATGGGAGAGGGAGCGGACG 780  
 Db 1427 TTCATGAAAGAAAGGGGTGAGAGATGGGGGAGCATGGGATGGGAGAGGGAGCGGACG 1486  
 Qy 781 GAGTTTCTGACGCGGACTTCTCGAGACGTACGAGCGGTACCAACGAGAGCCTGACG 840  
 Db 1487 GAGTTTCTGACGCGGACTTCTCGAGACGTACGAGCGGTACCAACGAGAGCCTGACG 1546  
 Qy 841 AACATGAGCAAGAGAGCTTCAAGAGATCTGGAATCTGGAAGTGGCTCTGCGCGC 900  
 Db 1547 AACATGAGCAAGAGAGCTTCAAGAGATCTGGAATCTGGAAGTGGCTCTGCGCGC 1606  
 Qy 901 ATGGAGAGCAAGAAACAACCGGCGTGGGCTGGAGAGCAACCGGCTGGGCGGACGACGCG 960  
 Db 1607 ATGGAGAGCAAGAAACAACCGGCGTGGGCTGGAGAGCAACCGGCTGGGCGGACGACGCG 1666  
 Qy 961 CGTGTGCGGAGCTGAGAGCTGAGAGCTGAGAGCTGAGAGAGAGAGAGAGAGAGAGAG 1020  
 Db 1667 CGTGTGCGGAGCTGAGAGCTGAGAGCTGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1726  
 Qy 1021 ACCGAGAGCAAGCTGACCGGCGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1080  
 Db 1727 ACCGAGAGCAAGCTGACCGGCGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1786  
 RESULT 3  
 US-10-205-823-171  
 ; Sequence 171, Application US/10205823  
 ; Publication No. US20030108963A1  
 ; GENERAL INFORMATION:  
 ; APPLICANT: Schlegel, Robert  
 ; APPLICANT: Monahan, John E.  
 ; APPLICANT: Endege, Wilson O.  
 ; APPLICANT: Ganavarapu, Manjula  
 ; APPLICANT: Gorbacheva, Bella  
 ; APPLICANT: Hoersch, Sebastian  
 ; APPLICANT: Kamatkar, Shubhangi  
 ; APPLICANT: Womsey, Angela M.  
 ; APPLICANT: Glatt, Karen  
 ; APPLICANT: Zhao, Xumei  
 ; APPLICANT: Anderson, Dustin  
 ; TITLE OF INVENTION: NOVEL GENES, COMPOSITIONS, KITS, AND  
 ; TITLE OF INVENTION: METHODS FOR IDENTIFICATION, ASSESSMENT, PREVENTION, AND  
 ; TITLE OF INVENTION: THERAPY OF PROSTATE CANCER  
 ; FILE REFERENCE: MRI-044  
 ; CURRENT APPLICATION NUMBER: US/10/205,823  
 ; PRIOR FILING DATE: 2002-07-25  
 ; PRIOR APPLICATION NUMBER: 60/307,982  
 ; PRIOR FILING DATE: 2001-07-25  
 ; PRIOR APPLICATION NUMBER: 60/314,356  
 ; PRIOR FILING DATE: 2001-08-22  
 ; PRIOR APPLICATION NUMBER: 60/325,020  
 ; PRIOR FILING DATE: 2001-09-25  
 ; PRIOR APPLICATION NUMBER: 60/341,746  
 ; PRIOR FILING DATE: 2001-12-12  
 ; PRIOR APPLICATION NUMBER: 60/362,158  
 ; PRIOR FILING DATE: 2002-03-05  
 ; NUMBER OF SEQ ID NOS: 455

; SOFTWARE: FastSeq for Windows Version 4.0  
 ; SEQ ID NO 171  
 ; LENGTH: 3624  
 ; TYPE: DNA  
 ; ORGANISM: Homo sapiens  
 US-10-205-823-171  
 Query Match 99.7%; Score 1076.8; DB 15; Length 3624;  
 Best Local Similarity 99.8%; Pred. No. 3.9e-268;  
 Matches 1078; Conservative 0; Mismatches 2; Indels 0; Gaps 0;  
 1 ATGGCCGAGCCATTCTTGTGAGAAATATCAACACCGCTCAAACTAGCAACTGTACAGT 60  
 Db 630 ATGGCCGAGCCATTCTTGTGAGAAATATCAACACCGCTCAAACTAGCAACTGTACAGT 749  
 Qy 61 GCTGCTGCTGTCAGAGAGAGCTGAACCTGAGCCGCCCGGAGGCGGAGAGAGAGAGG 120  
 Db 750 GCTGCTGCTGTCAGAGAGAGCTGAACCTGAGCCGCCCGGAGGCGGAGAGAGAGAGG 809  
 Qy 121 CCGAGAGAGACAGTAGGTGGCAATCGAGAGCGTTCCCGCAGTTGGTGGCCGTCCGCGG 180  
 Db 810 CCGAGAGAGACAGTAGGTGGCAATCGAGAGCGTTCCCGCAGTTGGTGGCCGTCCGCGG 869  
 Qy 181 CCGAGAGAGAGAGAGAGCTGGAATCCCACTCCCTTGCAGACCCAGGCTGTCCA 240  
 Db 870 CCGAGAGAGAGAGAGAGCTGGAATCCCACTCCCTTGCAGACCCAGGCTGTCCA 929  
 Qy 241 GAATCTAGCTGCTGAG 300  
 Db 930 GAATCTAGCTGCTGAG 989  
 Qy 301 GCGGACTTCCGCGCGCGCAG 360  
 Db 990 GCGGACTTCCGCGCGCGCAG 1049  
 Qy 361 CTTGTCTACTGCTCGAG 420  
 Db 1050 CTTGTCTACTGCTCGAG 1109  
 Qy 421 TGGGAGACAGCAGAGACAGCTGGGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 480  
 Db 1110 TGGGAGACAGCAGAGACAGCTGGGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1169  
 Qy 481 AAGCGCATTTGGAACCGTACTACAAAGCTGAAGCTGGAAGAGAGAGAGAGAGAGAGAG 540  
 Db 1170 AAGCGCATTTGGAACCGTACTACAAAGCTGAAGCTGGAAGAGAGAGAGAGAGAGAGAG 1229  
 Qy 541 AAACAGAGCCTTGAAGCTTCAAGATTCGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 600  
 Db 1230 AAACAGAGCCTTGAAGCTTCAAGATTCGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1289  
 Qy 601 GGGCCCTATACACCAAGCAGTTCCTCATGATGATCAAGACCAAGAGAGAGAGAGAGAG 660  
 Db 1290 GGGCCCTATACACCAAGCAGTTCCTCATGATGATCAAGACCAAGAGAGAGAGAGAGAG 1349  
 Qy 661 AAAACCGGCTTCTACTCCCAAGCGCGCGCGCCCAATCCGACGACACCGAGCATGACGAC 720  
 Db 1350 AAAACCGGCTTCTACTCCCAAGCGCGCGCGCCCAATCCGACGACACCGAGCATGACGAC 1409  
 Qy 721 TTCATGAAAGAAAGGGGTGAGAGATGGGGGAGCATGGGATGGGAGAGGGAGCGGACG 780  
 Db 1410 TTCATGAAAGAAAGGGGTGAGAGATGGGGGAGCATGGGATGGGAGAGGGAGCGGACG 1469  
 Qy 781 GAGTTTCTGACGCGGACTTCTCGAGACGTACGAGCGGTACCAACGAGAGCCTGACG 840  
 Db 1470 GAGTTTCTGACGCGGACTTCTCGAGACGTACGAGCGGTACCAACGAGAGCCTGACG 1529  
 Qy 841 AACATGAGCAAGAGAGCTTCAAGAGATCTGGAATCTGGAAGTGGCTCTGCGCGC 900  
 Db 1530 AACATGAGCAAGAGAGCTTCAAGAGATCTGGAATCTGGAAGTGGCTCTGCGCGC 1589  
 Qy 901 ATGGAGAGCAAGAAACAACCGGCGTGGGCTGGAGAGCAACCGGCTGGTGGCGACGACGCG 960

Db 1590 ATGAGACGAGAACACCGGCTGGGAGAGACGAGCGCTGGGCGACGACGCG 1649  
 Qy 961 CGTGTCGGGAGCTGGAGCTGAGCTGSAACCGGCTCGGCGCCGAGAACCTCCAGCTGCTG 1020  
 Db 1650 CGTGTCGGGAGCTGGAGCTGAGCTGSAACCGGCTCGGCGCCGAGAACCTCCAGCTGCTG 1709  
 Qy 1021 ACCGAGACGAACTGCAACCGGCGAGAGAGCGAGCGCCGCTTCCAAATTGGAGACTAG 1080  
 Db 1710 ACCGAGACGAACTGCAACCGGCGAGAGAGCGAGCGCCGCTTCCAAATTGGAGACTAG 1769

# RESULT 4 US-10-264-237-629/c

; Sequence 629, Application US/10264237  
 ; Publication No. US20040009491A1  
 ; GENERAL INFORMATION:  
 ; APPLICANT: Barse et al.  
 ; TITLE OF INVENTION: Nucleic Acids, Proteins, and Antibodies  
 ; FILE REFERENCE: P413P1  
 ; CURRENT APPLICATION NUMBER: US/10/264,237  
 ; PRIORITY FILING DATE: 2002-10-04  
 ; PRIORITY APPLICATION NUMBER: PCT/US01/16450  
 ; PRIORITY FILING DATE: 2001-05-18  
 ; PRIOR APPLICATION NUMBER: US 60/205,515  
 ; NUMBER OF SEQ ID NOS: 2876  
 ; SOFTWARE: Patent In Ver. 3.1  
 ; SEQ ID NO 629  
 ; LENGTH: 1083  
 ; TYPE: DNA  
 ; ORGANISM: Homo sapiens  
 ; FEATURE:  
 ; NAME/KEY: misc feature  
 ; LOCATION: (16)..(16)  
 ; OTHER INFORMATION: n equals a,t,g, or c  
 ; FEATURE:  
 ; NAME/KEY: misc feature  
 ; LOCATION: (27)..(27)  
 ; OTHER INFORMATION: n equals a,t,g, or c  
 ; FEATURE:  
 ; NAME/KEY: misc feature  
 ; LOCATION: (32)..(32)  
 ; OTHER INFORMATION: n equals a,t,g, or c  
 ; FEATURE:  
 ; NAME/KEY: misc feature  
 ; LOCATION: (111)..(111)  
 ; OTHER INFORMATION: n equals a,t,g, or c  
 ; US-10-264-237-629

Query Match 40.2%; Score 434.2; DB 12; Length 1083;  
 Best Local Similarity 98.9%; Pred. No. 2,6e-102;

Matches 444; Conservative 3; Mismatches 1; Indels 1; Gaps 1;

Qy 632 ATGATACGACGACGAGAGCCCGGATCTCAAAACCGGCTGTACTCAACGGGCGCGCG 691  
 Db 1083 ATGATACGACGACGAGAGCCCGGATCTCAAAACCGGCTGTACTCAACGGGCGCGCG 1024  
 Qy 692 CCAATCCGACGACGACGAGATGAGACTTCAATGAGAGAGGGGTTGAGAGAGATGGG 751  
 Db 1023 -CAATCCGACGACGACGAGATGAGACTTCAATGAGAGAGGGGTTGAGAGAGATGGG 965  
 Qy 752 GCAGCATGGATGAGAGGAGGAGCGGAGGATTTCTGAGCGGAGCTTCTCGAGAGCT 811  
 Db 964 GCAGCATGGATGAGAGGAGGAGGAGCGGAGGATTTCTGAGCGGAGCTTCTCGAGAGCT 905  
 Qy 812 ACAGCGGTTACCAACGAGAGGCTTCGAGAACATGAGCAAGAGAGCTCATAGAGAGT 871  
 Db 904 ACAGCGGTTACCAACGAGAGGCTTCGAGAACATGAGCAAGAGAGCTCATAGAGAGT 845  
 Qy 872 ACCTGGAACGAGAGAGTCTCTCGGAGATGAGAGACGAGAACACCGGCTGGCGCTG 931  
 Db 844 ACCTGGAACGAGAGAGTCTCTCGGAGATGAGAGACGAGAACACCGGCTGGCGCTG 785

Qy 932 AAGACAGCGCTGGGTGGCGACGACGCGCTGTGCGGAGCTGGAGCTGGAGCTGGACC 991  
 Db 784 AAGACAGCGCTGGGTGGCGACGACGCGCTGTGCGGAGCTGGAGCTGGAGCTGGACC 725  
 Qy 992 GGCTGGCGCGGAGAACCTCCAGCTGCTGACCGAGAACGAACTGACCGGCGAGAGAGC 1051  
 Db 724 GGCTGGCGCGGAGAACCTCCAGCTGCTGACCGAGAACGAACTGACCGGCGAGAGAGC 665  
 Qy 1052 GAGCGCGCTTCCAAATTGGAGACTAG 1080  
 Db 664 GAGCGCGCTTCCAAATTGGAGACTAG 636

# RESULT 5 US-09-918-995-403/c

; Sequence 403, Application US/09918995  
 ; Publication No. US20030073623A1  
 ; GENERAL INFORMATION:  
 ; APPLICANT: Hyseq, Inc.  
 ; TITLE OF INVENTION: NOVEL NUCLEIC ACID SEQUENCES OBTAINED  
 ; FILE REFERENCE: FROM VARIOUS CDNA LIBRARIES  
 ; CURRENT APPLICATION NUMBER: US/09/918,995  
 ; CURRENT FILING DATE: 2001-07-30  
 ; PRIOR APPLICATION NUMBER: US/09/235,076  
 ; PRIORITY FILING DATE: 1999-01-20  
 ; NUMBER OF SEQ ID NOS: 38054  
 ; SOFTWARE: FastSeq for Windows Version 3.0  
 ; SEQ ID NO 403  
 ; LENGTH: 495  
 ; TYPE: DNA  
 ; ORGANISM: Homo sapiens  
 ; FEATURE:  
 ; NAME/KEY: misc feature  
 ; LOCATION: (1)..(495)  
 ; OTHER INFORMATION: n = A,T,C or G  
 ; US-09-918-995-403

Query Match 34.1%; Score 368.8; DB 11; Length 495;  
 Best Local Similarity 99.5%; Pred. No. 1.7e-85;

Matches 370; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

Qy 709 AGCGATGACGACTTATGAGAGAGGGGTTGAGAGAGATGGGGGACGATGGATGGGA 768  
 Db 495 AGCGATGACGACTTATGAGAGAGGGGTTGAGAGAGATGGGGGACGATGGATGGGA 436  
 Qy 769 GGGAGCGGAGGAGATTTCTGAGCGGAGCTTCTCGAGAGACCTAAGACGGTACACACG 828  
 Db 435 GGGAGCGGAGGAGATTTCTGAGCGGAGCTTCTCGAGAGACCTAAGACGGTACACACG 376  
 Qy 829 GAGAGCTTCAGAACATGAGCAAGAGAGGCTCATCAAGAGTACCTGGAACTGGAGAG 888  
 Db 375 GAGAGCTTCAGAACATGAGCAAGAGAGGCTCATCAAGAGTACCTGGAACTGGAGAG 316  
 Qy 889 TGCCTCTGCGGATGAGAGAGCAACACCGGCTTGGGAGAGCAAGCGGCTGGGT 948  
 Db 315 TGCCTCTGCGGATGAGAGAGCAACACCGGCTTGGGAGAGCAAGCGGCTGGGT 256  
 Qy 949 GGCAGAGAGCGCGGTGTGCGGAGCTGAGAGCTGAGAGCTGAGACCGGCTGGCGCCGAGAC 1008  
 Db 255 GGCAGAGAGCGCGGTGTGCGGAGCTGAGAGCTGAGAGCTGAGACCGGCTGGCGCCGAGAC 196  
 Qy 1009 CTCAGCTGCTGACGAGAACGAACTGCAACCGGCGAGAGAGCGAGCGCGCTTCCAAAG 1068  
 Db 195 CTCAGCTGCTGACGAGAACGAACTGCAACCGGCGAGAGAGCGAGCGCGCTTCCAAAG 136  
 Qy 1069 TTTGAGACTAG 1080  
 Db 135 TTTGAGACTAG 124

RESULT 6  
 US-09-983-965-4895

```

Sequence 4895 Application US/03983965
Patent No. US20020137160A1
GENERAL INFORMATION:
APPLICANT: Warren, Wesley C.
APPLICANT: Tao, Nengbing
APPLICANT: Byatt, John C.
APPLICANT: Mathialagan, Nagappan
TITLE OF INVENTION: NUCLEIC ACID AND OTHER MOLECULES ASSOCIATED WITH LACTATION AND
FILE REFERENCE: 37-21(10297)C
CURRENT APPLICATION NUMBER: US/09/983.965
CURRENT FILING DATE: 2001-10-26
PRIOR APPLICATION NUMBER: US 09/465,231
PRIOR FILING DATE: 1999-12-15
PRIOR APPLICATION NUMBER: US 60/113,678
PRIOR FILING DATE: 1998-12-17
NUMBER OF SEQ ID NOS: 5912
SEQ ID NO 4895
LENGTH: 414
TYPE: DNA
ORGANISM: Bos taurus
FEATURE:
OTHER INFORMATION: Clone ID: 24-LIB34-008-Q1-E1-F7
US-09-983-965-4895

Query Match          33.6%; Score 362.8; DB 10; Length 414;
Best Local Similarity 92.3%; Pred. No. 6e-84;
Matches 382; Conservative 0; Mismatches 32; Indels 0; Gaps 0;

QY 462 GAGAGCGCCCGCCCAAGAGAAAGCGCATTTGGAAACCGACTACACAGCTGAACTGGAGAGA 521
DB 1 GAGAGCGCCCGCTCCCAAGAGAAAGCGCGCTTTGGAAACCGACTACATACGCTGACCTGGAGAGA 60
QY 522 GAAGAAAAAGTTCCAGCAGAAACAGAGCCTTCGAGCTTCAAGATCCGAGCCGAGATGTT 581
DB 61 GAAGAAAAAGTTCCATGATAAAAAAGAGCCTCGAGCTTCGAGATTCGAGCCGAGATGTT 120
QY 582 CGCCAAAGGCGCAGCGCGGTGCGCCCTTATACACACACAGCATGTTCTCATGATGATCACGA 641
DB 121 CGCCAAAGGCGCAGCGAGTGTCTCCCTATACACACAGCATGTTCTCATGATGATCACGA 180
QY 642 CCAGAGAGAGCGGATCTCAAAACCGGCTGTACTCCAAAGGCGGCGCGCGCAATCCGA 701
DB 181 CCAGAGAGAGCGGATCTTAAACCGGCTTATATCCCAAAAGGCGCGCTGCCAATCCGA 240
QY 702 CGACACGACGATGACGATTCATGSAAGAGGCGGTAGAGAGATGGGCGCACGATGG 761
DB 241 CGACACGACGATGATGATCTTATGAAAGAGCGGCGAGAGATGGGCGCACGACGCG 300
QY 762 GATGGAGAGGAGCGGCGAGCGAGTTTCTGACGCGGAGCTTCTGAGAGAGTACGAGCGGTA 821
DB 301 GATGGAGAGAGAGCGGCGAGCGAGTTTCTGACGCGGAGCTTCTGAGAGAGTATGAGCGGTA 360
QY 822 CCAACGAGAGAGCTCTGACGAACATGAGACACAGAGAGCTATCAAGAGATGACT 875
DB 361 CCAGCGGAGAGAGCTCTGACGAACATGAGACACAGAGAGCTATCAAGAGATGACT 414

RESULT 7
US-09-954-531-813
Sequence 813, Application US/09954531
Patent No. US20020165180A1
GENERAL INFORMATION:
APPLICANT: Weaver, Zoe
TITLE OF INVENTION: Process for Identifying Anti-Cancer Therapeutic Agents Using Can
FILE REFERENCE: 689290-77
CURRENT APPLICATION NUMBER: US/09/954.531
CURRENT FILING DATE: 2002-05-02
PRIOR APPLICATION NUMBER: US/60/233,133
PRIOR FILING DATE: 2000-09-18
PRIOR APPLICATION NUMBER: US/60/234,009
PRIOR FILING DATE: 2000-09-20

```

Query	Match	32.9%	Score 355.2	DB 10	Length 461
Db	72	ATGGCCGAGCATTCTTGTGACAAATATCAACACGAGCCTCAAACTAGCACTGTACAGT	60		
Qy	61	GCTGCTGCTGTCCAGAAAGCTGAACCTCTGAGGCCCCCCAGGCGCGAGAGCGGGTG	120		
Db	132	GCTGCTGCTGTCCAGAAAGCTGAACCTCTGAGGCCCCCCAGGCGGACCGAGGGGGTG	191		
Qy	121	CCGAGGAGGACAGTAGTGGCAATCGAGAGCGTTCCCGGAGTTGGGTGGCGGTCCGGGG	180		
Db	192	CCGAGGAGGACAGTAGTGGCAATCGAGAGCGTTCCCGGAGTTGGGTGGCGGTCCGGGG	251		
Qy	181	CCGAGGAGGAGGAGGAGCCTGGAATCCCAACCACTCCCTTGACAGCCAGGCTGTCCA	240		
Db	252	CCGAGGAGGAGGAGGAGCCTGGAATCCCAACCACTCCCTTGACAGCCAGGCTGTCCA	311		
Qy	241	GAATCTAGCTGCTGAGAGAGGGCGAGAGGGCCAGAAATGGGGACGACTCGTCCGCTGGC	300		
Db	312	GAATCTAGCTGCTGAGAGAGGGCGAGAGGGCCAGAAATGGGGACGACTCGTCCGCTGGC	371		
Qy	301	GGGCACTTCCCGCGCGCGCGGACAGTAGTGAACCGACGCCCGGAGGCGGAGCTGTGCGCCAG	360		
Db	372	GGGCACTTCCCGCGCGCGCGGACAGTAGTGAACCGACGCCCGGAGGCGGAGCTGTGCGCCAG	430		
Qy	361	CCTTGTCATGACTCCGAGGCCAGTAAGTTGG	391		
Db	431	CCTTGTCATGACTCCGAGGCCAGTAAGTTGG	461		

```

; APPLICANT: MASUHO, YASUHIKO
; TITLE OF INVENTION: NOVEL FULL-LENGTH CDNA
; FILE REFERENCE: 084335/0160
; CURRENT APPLICATION NUMBER: US/10/094,749
; CURRENT FILING DATE: 2002-03-12
; PRIOR APPLICATION NUMBER: 60/350,435
; PRIOR FILING DATE: 2002-01-24
; PRIOR APPLICATION NUMBER: JP 2001-328381
; PRIOR FILING DATE: 2001-09-14
; NUMBER OF SEQ ID NOS: 3381
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 1223
; LENGTH: 1330
; TYPE: DNA
; ORGANISM: Homo sapiens
US-10-094-749-1223

Query Match      16.2%; Score 174.8; DB 13; Length 1330;
Best Local Similarity 60.5%; Pred. No. 2.6e-35;
Matches 362; Conservative 0; Mismatches 197; Indels 39; Gaps 3;

QY 441 GCTGGGAGAAAAAATATAGAGACGCCCGTCCAGAAAGACGGCATTGGAACCGTA 500
    |||||
DB 494 GCTGGCCCGAAGAAACACCGTGGCGGCATGAAAGCGAAGCACTGGGACCTTA 553
    |||||
QY 501 CTACAACTGAATCTGGAGAGAGAAAGAAAGTTGACGAGAAACAGAGCCTTGAGCTTC 560
    |||||
DB 554 CTTGAGAGCTGAGCTGGGCTGAGAAACACAGCGGATGAGAGGACAGAGGAGGCGCTC 613
    |||||
QY 561 AAGGATCCGAGCCGAGATGTTCCCAAGGCGCCGCTGCGCCCTATTAACCAACGCA 620
    |||||
DB 614 CCGGGTCCGGAAGATGTTCCCAAGGCGCCGCTGCGCCCTATTAACCAACCA 673
    |||||
QY 621 GTTCCTCATGATGATCAAGACAGAGAGAGCGGATCTCAAAACCGCCTGATCTCAA 680
    |||||
DB 674 GTTCTGATGATGATGACAGAGAGAGCGGAGAGCC-----CAACTTGATGTGCCCCAT 725
    |||||
QY 681 GCGGGCCCGCCCAATTCGACGACACACGAGTACGACTTCAGAAAGAGGGGTGA 740
    |||||
DB 726 GGGATCTCCACCCAGGTTCCAG-----TGGGAGAG 757
    |||||
QY 741 GAGAGATGGGGGCGAGATGGGATGGGAGGAGACGACAGACTTCTGACGGGAGATT 800
    |||||
DB 758 TAGAGCCGGGAGAGATGATGGGCGGCGGCGACGAGATTCTCAGGAGAAAGACTT 817
    |||||
QY 801 CTCGAGAGCTAGACGAGTACCAACAGAGAGCTGCAAGATAGACAGAGAGCT 860
    |||||
DB 818 CTCTGAGACTTACGAGAGCTTCCACACGAGAGCTGCAAGGCGGACAGAGAGAGCT 877
    |||||
QY 861 CATCAAGAGTACCTGGAACCTGGAAGTCTCTGCGCATGAGAGAGAGCAACACCG 920
    |||||
DB 878 GGTGCGAGACTACTGAGAGCTGGAAGAGCGCTGTGCGAGGCGAGAGAGACTAGAG 937
    |||||
QY 921 GCTGCGGCTGAGAGAGCAAGCGCTGGGTGGCGA---CGAGCGCGTGTGCGGAGCTTGA 977
    |||||
DB 938 GCTGCGAGAGCTGAGAGCGCTGCAACCGGACGAGCTGCTGCGGCAAGTGAAGAGTGGC 997
    |||||
QY 978 GCTGAGAGCTGAGACCGGCTGGCGCGGAGAACTCCAGCTGCTGACGAGAGAACTG 1035
    |||||
DB 998 TGCCGAGGTCAGAGGCTCGAGACCGAAGAACAGCGGCTTCTGACAGAGAACAGATG 1055
    |||||

RESULT 9
US-10-108-260A-1496/c
; Sequence 1496, Application US/10108260A
; Publication No. US20040005560A1
; GENERAL INFORMATION:
; APPLICANT: HELIX RESEARCH INSTITUTE
; TITLE OF INVENTION: NO. US20040005560A1e1 full length cDNA
; FILE REFERENCE: H1-A0106
; CURRENT APPLICATION NUMBER: US/10/108,260A
; CURRENT FILING DATE: 2002-03-27
; NUMBER OF SEQ ID NOS: 5458

```

```

; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 1496
; LENGTH: 2048
; TYPE: DNA
; ORGANISM: Homo sapiens
US-10-108-260A-1496

Query Match      16.0%; Score 173.2; DB 12; Length 2048;
Best Local Similarity 60.4%; Pred. No. 7.2e-35;
Matches 361; Conservative 0; Mismatches 198; Indels 39; Gaps 3;

QY 441 GCTGGGAGAAAAAATATAGAGACGCCCGTCCAGAAAGACGGCATTGGAACCGTA 500
    |||||
DB 1650 GCTGGCCCGAAGAAACACCGTGGCGGCATGAAAGCGAAGCACTGGGACCTTA 1591
    |||||
QY 501 CTACAACTGAATCTGGAGAGAGAAAGAAAGTTGACGAGAAACAGAGCCTTGAGCTTC 560
    |||||
DB 1590 CTTGAGAGCTGAGCTGGGCTGAGAAACACAGCGGATGAGAGGACAGAGGCGCTC 1531
    |||||
QY 561 AAGGATCCGAGCCGAGATGTTGCCAAGGCGCAGCGGCTGCGCCCTATTAACCAACGCA 620
    |||||
DB 1530 CCGGGTCCGGAAGATGTTGCCAAGGCGCAGCGGCTGCGCCCTATTAACCAACCA 1471
    |||||
QY 621 GTTCCTCATGATGATCAAGACAGAGAGAGCGGATCTCAAAACCGCCTGATCTCAA 680
    |||||
DB 1470 GTTCTGATGATGATGACAGAGAGAGCGGAGAGCC-----CAACTTGATGTGCCCCAT 1419
    |||||
QY 681 GCGGGCCCGCCCAATTCGACGACACACGAGTACGACTTCATGAGAAAGGGGTGA 740
    |||||
DB 1418 GGGATCTCCACCCAGGTTCCAG-----TGGGAGAG 1387
    |||||
QY 741 GAGAGATGGGGGCGAGATGGGATGGAGAGGAGACGACGAGTTCTGACGGGACTT 800
    |||||
DB 1386 TAGAGCCGGGAGACATGATGGGCGGCGGCGAGCGACGATGATGATTCAGCGGAGACTT 1327
    |||||
QY 801 CTCGAGAGCTAGACGAGTACCAACAGAGAGCTGCAAGATAGAGAACAGAGCT 860
    |||||
DB 1326 CTCTGAGACTTACGAGAGCTTCCACACGAGAGCTGCAAGGCGGCGCAGACAGAGCT 1267
    |||||
QY 861 CATCAAGAGTACCTGGAACCTGGAAGTCTCTGCGCATGAGAGAGCAACACCG 920
    |||||
DB 1266 GGTGCGAGACTACTGAGAGCTGGAAGAGCGCTGTGCGAGGAGAGAGACTAGAG 1207
    |||||
QY 921 GCTGCGGCTGAGAGCAAGCGGCTGGGTGGCGA---CGAGCGCGTGTGCGGAGCTTGA 977
    |||||
DB 1206 GCTGCGAGAGCTGAGAGCGGCTGCAACCGGACGAGCTGCTGCCACAGTGAAGAGCTGGC 1147
    |||||
QY 978 GCTGAGAGCTGAGACCGGCTGGCGCGGAGAACTCCAGCTGCTGACGAGAGAACTG 1035
    |||||
DB 1146 TGCCGAGGTCAGAGGCTCCGAGACCGAAGAACAGCGGCTTCTGACAGAGAACAGATG 1089
    |||||

RESULT 10
US-10-294-804-3
; Sequence 3, Application US/10294804
; Publication No. US20030133948A1
; GENERAL INFORMATION:
; APPLICANT: Robertson, Eric S.
; APPLICANT: Corbett, Murray A.
; TITLE OF INVENTION: Methods to Inhibit or Enhance the Binding of Viral DNA
; FILE REFERENCE: UM-03778
; CURRENT APPLICATION NUMBER: US/10/294,804
; CURRENT FILING DATE: 2002-11-14
; PRIOR APPLICATION NUMBER: US/09/410,399
; PRIOR FILING DATE: 1999-10-01
; NUMBER OF SEQ ID NOS: 6
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 3
; LENGTH: 1926
; TYPE: DNA
; ORGANISM: Epstein-Barr virus
US-10-294-804-3

```









Db 1274 GAGGGGACAGAGGGGACAGAGGAGGGGACAGAGGGGACAGAGGAGGGGACAGAG 1333  
Qy 1006 AACCTCCAGCTGCTGACCGAGAACTGACCGGACAGAGGAGCGCG 1059  
Db 1334 GGGCAGAGACAGAGAGGGGACAGAGGAGGACAGAGAGGGGACAGAGCAG 1387

## RESULT 14

US-10-050-898-283/c  
Sequence 283, Application US/10050898  
Publication No. US20030175711A1

## GENERAL INFORMATION:

APPLICANT: Renner, Wolfgang A.  
APPLICANT: Bachmann, Martin  
APPLICANT: Tissot, Alain  
APPLICANT: Maurer, Patrick  
APPLICANT: Lechner, Franziska  
APPLICANT: Seibel, Peter  
APPLICANT: Ploesek, Christine  
APPLICANT: Ortman, Rainer  
APPLICANT: Luond, Rainer  
APPLICANT: Staufenbiel, Matthias  
APPLICANT: Frey, Peter

TITLE OF INVENTION: Molecular Antigen Array

FILE REFERENCE: 1700.0190005

CURRENT APPLICATION NUMBER: US/10/050,898

PRIOR FILING DATE: 2002-01-18

PRIOR APPLICATION NUMBER: US 60/262,379

PRIOR FILING DATE: 2001-01-19

PRIOR APPLICATION NUMBER: US 60/288,549

PRIOR FILING DATE: 2001-05-04

PRIOR APPLICATION NUMBER: US 60/326,998

PRIOR FILING DATE: 2001-10-05

PRIOR APPLICATION NUMBER: US 60/331,045

PRIOR FILING DATE: 2001-11-07

NUMBER OF SEQ ID NOS: 350

SOFTWARE: PatentIn Ver. 2.1

SEQ ID NO 283

LENGTH: 10233

TYPE: DNA

ORGANISM: Artificial Sequence

FEATURE:

OTHER INFORMATION: pcep-Xa-Fc construct

NAME/KEY: misc.feature

LOCATION: (9820)..(9820)

OTHER INFORMATION: n is a, c, g, or t

US-10-050-898-283

Query Match 6.3%; Score 68.4; DB 13; Length 10233;  
Best Local Similarity 44.0%; Pred. No. 1.1e-07;

Matches 288; Conservative 0; Mismatches 366; Indels 0; Gaps 0;

Qy 406 GGGGGCAGAGAGAGTGGGACAGACAGAGAGAGAGTGGGAGAGAGAGAGAGAGAGAGAG 465  
Db 4191 GGGCAG 4132  
Qy 466 CGCCGCTCCAG 525  
Db 4131 GAGGGGACAG 4072  
Qy 526 AAAAGTTCAG 585  
Db 4071 CAG 4012  
Qy 586 AAGGCGCAG 645  
Db 4011 GGGCAG 3952  
Qy 646 GAGGAG 705  
Db 3951 GAGGGGACAG 3892

Qy 706 ACCAGCAGTACGACTTCATGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 765  
Db 3891 CAG 3832  
Qy 766 GAGGAG 825  
Db 3831 GAGGGGACAG 3772  
Qy 826 ACGAG 885  
Db 3771 CAG 3712  
Qy 886 AAGTGCCTCTCGCGCATGAG 945  
Db 3711 GAGCAG 3652  
Qy 946 GGTGCGACAG 1005  
Db 3651 GAGGGGACAG 3592  
Qy 1006 AACCTCCAGCTGCTGACCGAGAACTGCAACCGGACAGAGAGAGAGAGAGAGAGAG 1059  
Db 3591 GGGCAGAGACAG 3538

## RESULT 15

US-10-050-902-283/c

Sequence 283, Application US/10050902

Publication No. US20030175290A1

## GENERAL INFORMATION:

APPLICANT: Renner, Wolfgang A.  
APPLICANT: Bachmann, Martin  
APPLICANT: Tissot, Alain  
APPLICANT: Maurer, Patrick  
APPLICANT: Lechner, Franziska  
APPLICANT: Seibel, Peter  
APPLICANT: Ploesek, Christine  
APPLICANT: Ortman, Rainer  
APPLICANT: Luond, Rainer  
APPLICANT: Staufenbiel, Matthias  
APPLICANT: Frey, Peter

TITLE OF INVENTION: Molecular Antigen Array

FILE REFERENCE: 1700.0190004

CURRENT APPLICATION NUMBER: US/10/050,902

PRIOR FILING DATE: 2002-01-18

PRIOR APPLICATION NUMBER: US 60/262,379

PRIOR FILING DATE: 2001-01-19

PRIOR APPLICATION NUMBER: US 60/288,549

PRIOR FILING DATE: 2001-05-04

PRIOR APPLICATION NUMBER: US 60/326,998

PRIOR FILING DATE: 2001-10-05

PRIOR APPLICATION NUMBER: US 60/331,045

PRIOR FILING DATE: 2001-11-07

NUMBER OF SEQ ID NOS: 350

SOFTWARE: PatentIn Ver. 2.1

SEQ ID NO 283

LENGTH: 10285

TYPE: DNA

ORGANISM: Artificial Sequence

FEATURE:

OTHER INFORMATION: pcep-Xa-Fc construct

NAME/KEY: misc.feature

LOCATION: (9872)..(9872)

OTHER INFORMATION: n is a, c, g, or t

US-10-050-902-283

Query Match 6.3%; Score 68.4; DB 13; Length 10285;  
Best Local Similarity 44.0%; Pred. No. 1.1e-07;

Matches 288; Conservative 0; Mismatches 366; Indels 0; Gaps 0;

Qy 406 GGGGGCAGAGAGAGTGGGACAGACAGAGAGAGAGTGGGAGAGAGAGAGAGAGAGAGAG 465  
Db 4191 GGGCAG 4132  
Qy 466 CGCCGCTCCAG 525

Db 4131 GAGGGGACAGACAGAGAGGGGCAAGACAGAGAGAGGGGCAAGAGGGGCAAGAGGGG 4072  
QY 526 AAAAAGTTGACAGAGAAACAGAGCCTTCAGAGCTTCAAGATCCGAGCCGAGATGTTGCGC 585  
Db 4071 CAGGAGCAGAGAGAGGGGCAAGAGCAGAGAGAGGGGCAAGAGGGGCAAGAGAGAG 4012  
QY 586 AAGGGCCAGCCGCTCCGCCCTTAACACACGCAATTCTCATGATGATCAAGACAG 645  
Db 4011 GGGCAGAGAGGGGCAAGAGGGGCAAGAGAGAGGGGCAAGAGAGAGAGAGGGGCAAG 3952  
QY 646 GAGGAGCCGGATCTCAAAACCGGCTGTACTCCAGGGGCGCGCCAAATCCGAGAC 705  
Db 3951 GAGGGGCAAGAGAGAGAGGGGCAAGAGGGGCAAGAGGGGCAAGAGAGAGAGAGGGG 3892  
QY 706 ACCAGCATGACGACTTCAATGAAAGAGGGGGTGAAGAGATGGGGGCAAGCCATGGGATG 765  
Db 3891 CAGGAGCAGAGAGGGGCAAGAGGGGCAAGAGGGGCAAGAGAGAGAGAGAGAGAGAG 3832  
QY 766 GAGGGGACGGCAGCGAGTTTCTGACCGGGACTTCTCGAGACGTACGAGCGGTACAC 825  
Db 3831 GAGGGGCAAGAGGGGCAAGAGGGGCAAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 3772  
QY 826 AAGGAGAGCCTGAGAACTGAGCAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 885  
Db 3771 CAGGAGGGGCAAGAGAGAGGGGCAAGAGGGGCAAGAGAGAGAGAGAGAGAGAGAG 3712  
QY 886 AAGTGCCTCTCGCGCATGAGAGAGAGAAACAACCGGCTGCGCTGAGAGCAAGCGGCTG 945  
Db 3711 GAGCAGAGAGGGGCAAGAGGGGCAAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 3652  
QY 946 GGTGGGACGACGCGCGGTGTGCGGGAGCTGAGAGCTGAGCTGGAACGGCTGCGCGCGAG 1005  
Db 3651 GAGGGGCAAGAGGGGCAAGAGAGAGGGGCAAGAGGGGCAAGAGAGAGAGAGAGAG 3592  
QY 1006 AACCTCAGAGCTGAGCCGAGAAAGCACTGCAACCGGAGAGAGAGAGAGAGAGAGAG 1059  
Db 3591 GGGCAGAGAGAGAGGGGCAAGAGGGGCAAGAGAGAGAGAGAGAGAGAGAGAGAGAG 3538

Search completed: February 5, 2004, 13:27:01  
Job time : 495 secs

GenCore version 5.1.6  
Copyright (c) 1993 - 2004 Compugen Ltd.

OM protein - nucleic search, using frame\_plus\_p2n model

Run on: February 5, 2004, 13:32:45 ; Search time 4658 Seconds  
(without alignments)  
3152.975 Million cell updates/sec

Title: US-09-972-758A-2

Perfect score: 1910  
Sequence: 1 NAEPLSEYQHQPOTSINCTG.....LTENELHROQERAPLSKFGD 359

Scoring table: BLOSUM62  
Xgapop 10.0 , Xgapext 0.5  
Ygapop 10.0 , Ygapext 0.5  
Fgapop 6.0 , Fgapext 7.0  
Delop 6.0 , Delext 7.0

Searched: 2888711 segs, 2045481386 residues  
Total number of hits satisfying chosen parameters: 5777422

Minimum DB seq length: 0  
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%  
Maximum Match 100%  
Listing first 45 summaries

Command line parameters:  
-MODE=frame\_plus\_p2n.model -DRV=xlh  
-Q=/cgn2.1/USPTO.spool/US09977358/runat\_05022004\_095003.1189/app\_query.fasta.1.519  
-DB=GenEmbl -QFMT=fastap -SUFFIX=type -MTNMATCH=0.1 -LOOPEL=0 -LOOPEXT=0  
-UNITS=b1ts -START=1 -END=1 -MATRIX=blsum62 -TRANS=human40.cdt -LIST=45  
-LOCALLEN=200 -THR SCORE=pct -THR MAX=100 -THR MIN=0 -ALIGN=15 -MODE=LOCAL  
-OUTFMT=pgc -NORM=ext -HEAPSIZE=500 -MINLEN=0 -MAXLEN=2000000000  
-USER=US09977358.@CNCN\_1.1.3508@runat\_05022004\_095003.1189 -NCPU=6 -ICPU=3  
-NO\_MAP -LANGUAGE=EN -NEG SCORES=0 -WAIT -DSPELOCK=100 -LONGLOG  
-DEV TIMEOUT=120 -WARN TIMEOUT=30 -THREADS=1 -XGAPOP=10 -XGAPEXT=0.5 -FGAPOP=6  
-FGAPEXT=7 -YGAPOP=10 -YGAPEXT=0.5 -DELOP=6 -DELEXT=7

Database : GenEmbl:.\*  
1: gb\_da.\*  
2: gb\_hcg.\*  
3: gb\_in.\*  
4: gb\_om.\*  
5: gb\_ov.\*  
6: gb\_pat.\*  
7: gb\_ph.\*  
8: gb\_pl.\*  
9: gb\_pr.\*  
10: gb\_ro.\*  
11: gb\_sts.\*  
12: gb\_sy.\*  
13: gb\_un.\*  
14: gb\_vl.\*  
15: em\_da.\*  
16: em\_fun.\*  
17: em\_hum.\*  
18: em\_in.\*  
19: em\_mu.\*  
20: em\_om.\*  
21: em\_or.\*  
22: em\_ov.\*  
23: em\_pat.\*  
24: em\_ph.\*  
25: em\_pl.\*  
26: em\_ro.\*  
27: em\_sts.\*  
28: em\_un.\*

29: em\_vi.\*  
30: em\_hcg\_hum.\*  
31: em\_hcg\_inv.\*  
32: em\_hcg\_other.\*  
33: em\_hcg\_mus.\*  
34: em\_hcg\_pln.\*  
35: em\_hcg\_rtd.\*  
36: em\_hcg\_mam.\*  
37: em\_hcg\_vrt.\*  
38: em\_sy.\*  
39: em\_hcg\_hum.\*  
40: em\_hcg\_mus.\*  
41: em\_hcg\_other.\*

Pred. No. is the number of results predicted by chance to have a  
score greater than or equal to the score of the result being printed,  
and is derived by analysis of the total score distribution.

#### SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	1910	100.0	2086	9 BC006460	BC006460 Homo sapi
2	1910	100.0	2199	6 BD106410	BD106410 Secreted
3	1910	100.0	3624	9 AB021179	AB021179 Homo sapi
4	1910	100.0	35050	9 AC142472	AC142472 Homo sapi
5	1910	100.0	38849	9 AC138150	AC138150 Homo sapi
6	1910	100.0	100836	2 AC087298	AC087298 Homo sapi
7	1903	99.6	2178	6 BD160090	BD160090 Primer to
8	1903	99.6	2178	9 AK023624	AK023624 Homo sapi
9	1858.5	97.3	142326	2 AC024047	AC024047 Homo sapi
10	1587.5	83.1	3402	10 AY090614	AY090614 Mus muscu
11	1587.5	83.1	131002	10 AL731805	AL731805 Mouse DNA
12	1584.5	82.6	213625	2 AC120950	AC120950 Rattus no
13	1577.5	82.6	1488	10 BC022111	BC022111 Mus muscu
14	875	45.8	111803	2 AC136172	AC136172 Rattus no
15	714	37.4	461	6 BD058159	BD058159 Secreted
16	629	32.9	523	6 AX331371	AX331371 Sequence
17	557.5	29.2	239464	2 AC103079	AC103079 Rattus no
18	557.5	29.2	267540	2 AC107153	AC107153 Rattus no
19	556.5	29.1	1330	6 AX714539	AX714539 Sequence
20	556.5	29.1	1330	9 AK056946	AK056946 Homo sapi
21	550	28.8	96875	9 AC008105	AC008105 Homo sapi
22	550	28.8	223442	10 AL662804	AL662804 Mouse DNA
23	545.5	28.6	1238	9 BC025970	BC025970 Homo sapi
24	545	28.5	2048	9 AK096785	AK096785 Homo sapi
25	378	19.8	1980	3 AK115840	AK115840 Clona int
26	333	17.4	225116	2 AC112538	AC112538 Rattus no
27	300	15.7	111803	2 AC136172	AC136172 Rattus no
28	245	12.8	342	6 BD059439	BD059439 Secreted
29	217	11.4	1561	3 BT001551	BT001551 Drosophila
30	214	11.2	1536	3 AY051786	AY051786 Drosophila
31	205.5	10.8	183439	3 AC007809	AC007809 Drosophila
32	205.5	10.8	232741	3 AB003705	AB003705 Drosophila
33	205.5	10.8	237119	2 AC017740	AC017740 Drosophila
34	170	8.9	9121	4 OATRICH	Z18092 O. cuniculus
35	165.5	8.7	9344	4 OATRICH	Z18361 O. aries tri
36	165	8.6	6644	4 RABWHCP	M7812 Rabbit myos
37	165	8.6	218667	5 AC095722	AC095722 Rattus no
38	164.5	8.6	4143	5 GSDYNACT	X62773 Gallus gall
39	162.5	8.5	89924	2 AC135973	AC135973 Homo sapi
40	162	8.5	4152	9 BCD49849	M81105 Homo sapien
41	162	8.5	4271	9 BCD49849	BCD49849 Homo sapi
42	162	8.5	7396	6 AX686226	AX686226 Sequence
43	162	8.5	278652	6 AC073809	AC073809 Mus muscu
44	162	8.5	314133	2 AC129914	AC129914 Homo sapi
45	160	8.4	205288	2 AC140253	AC140253 Mus muscu

RESULT 1

#### ALIGNMENTS

BC006460 2086 bp mRNA linear PRI 12-JUL-2001  
 LOCUS Homo sapiens, HMBA-inducible, clone MGC:1880 IMAGE:3535529, mRNA,  
 DEFINITION complete cds.  
 ACCESSION BC006460  
 VERSION BC006460.1 GI:13623668  
 KEYWORDS MGC.  
 SOURCE Homo sapiens (human)  
 ORGANISM Homo sapiens  
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.  
 1 (bases 1 to 2086)  
 Strausberg, R.  
 Direct Submision  
 Submitted (09-APR-2001) National Institutes of Health, Mammalian  
 Gene Collection (MGC), Cancer Genomics Office, National Cancer  
 Institute, 31 Center Drive, Room 11A03, Bethesda, MD 20892-2550,  
 USA  
 NIH-MGC Project URL: <http://mgc.ncl.nih.gov>  
 Contact: MGC help desk  
 Email: [cgabs-r@mail.nih.gov](mailto:cgabs-r@mail.nih.gov)  
 Tissue Procurement: DCTD/BTP  
 CDNA Library Preparation: Rubin Laboratory  
 CDNA Library Arrayed by: The I.M.A.G.E. Consortium (LNL)  
 DNA Sequencing by: Genome Sequence Centre,  
 BC Cancer Agency, Vancouver, BC, Canada  
 info@cgsc.bc.ca  
 Steven Jones, Jennifer Asano, Ian Bosdet, Yaron Buterfeld,  
 Susana Chan, Readman Chiu, Chris Fell, Erin Garland, Ran Guin,  
 Leticia Hsiao, Martin Krzywinski, Reta Kutsche, Oliver Lee, Soo  
 Sen Lee, Victor Ling, Carrie Mathewson, Candice McLeavy, Steven  
 Ness, Pawan Pandon, Anna-Lisa Prabhu, Parvaneh Saeedi, Jacqueline  
 Schein, Duane Smalish, Michael Smith, Lorraine Spence, Jeff Stolt,  
 Michael Thorne, Miranada Tsai, Natasja van den Bosch, Jill Vardy,  
 George Yang, Scott Zuyderduyn, Marco Marra.  
 Clone distribution: MGC clone distribution information can be found  
 through the I.M.A.G.E. Consortium/LNL at: <http://image.lnl.gov>  
 Series: IRAL Plate: 8 Row: d Column: 1  
 This clone was selected for full length sequencing because it  
 passed the following selection criteria: matched mRNA gi: 4062855.  
 Location/Qualifiers  
 1. 2086  
 /organism="Homo sapiens"  
 /mol\_type="mRNA"  
 /db\_xref="LocustID:10614"  
 /db\_xref="taxon:9606"  
 /clone="MGC:1880 IMAGE:3535529"  
 /issue\_type="Lung, small cell carcinoma"  
 /clone\_1b="NIH MGC 7"  
 /lab\_host="DH10B-R"  
 /note="Vector: pOTB7"  
 606..1685  
 /codon\_start=1  
 /product="HMBA-inducible"  
 /protein\_id="AAH06460.1"  
 /db\_xref="GI:13623668"  
 /translation="MAAPPLSYOHOPOTNCTGAAYOELNPPPPGAERVRP  
 SRNOSRAFPQUGRPGPEBEGSLESPPPIQTACESSSTRGEGNGDSSAGSD  
 FPPPAVEPTPEAEELAQPHOSEASKLAPAGGESEWQQQRQLGKKHRRPPEK  
 KRPKPYKLTWEKKKPFDEKOSLRASRIAEFAGQVAPYNTQFLMDHDQEP  
 DLKTGLYSKRAAKSDTDDDEMEGEGSDGSDGSDGSEFLQRPDSFYERHT  
 ESIQNSKQELIKYELKELKSRMEDENNRLLRSKRIGSGDAPARELELEDRRA  
 ENQLTELELHQGRAPLSKRGD"  
 BASE COUNT 510 a 505 c 618 g 453 t  
 ORIGIN  
 Alignment Scores:  
 Pred. No.: 1,27e-92 Length: 2086  
 Score: 1910.00 Matches: 359  
 Percent Similarity: 100.00% Conservative: 0  
 Best Local Similarity: 100.00% Mismatches: 0  
 Query Match: 100.00% Indels: 0

DB: 9 Gaps: 0  
 US-09-972-758a-2 (1-359) x BC006460 (1-2086)  
 QY 1 MetAAGLProPhenSerGluTyrGlnHsGlnProGlnThrsSerAsnGlyThrGly 20  
 DB 606 ATGGCCGAGCCATTCTTGTCAAGAAATTCACACCAAGCTCAACCTGACACTGTACAGGT 665  
 QY 21 AAlaAAlaValGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGln 40  
 DB 666 GGTGCTGCTGTCAAGAAAGCTGAACCTGTAGCGGCCCCCGAGCGCGGAGAGCGGGT 725  
 QY 41 ProGlnGlnAspSerArgTyrGlnSerArgAlaPheProGlnGlnGlnGlnGlnGlnGln 60  
 DB 726 CCGAGAGAGACAGTAAAGTGGAGCAATCGAAGAGGTTCCCAATTGGGTGGCGCTCGAGG 785  
 QY 61 ProGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGln 80  
 DB 786 CCGAG 845  
 QY 81 GluSerSerCysLeuArgGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGln 100  
 DB 846 GAATCTAGCTGCTGTAG 905  
 QY 101 GlnAspPheProProProAlaGlnValGlnProThrProGlnAlaGlnLeuAlaGln 120  
 DB 906 GGCAGCTTCCCGCCGCGCAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 965  
 QY 121 ProGlnHsAspSerGlnAlaSerGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGln 140  
 DB 966 CTTGTCAAGACTCCGAGGCCAGTAAAGTGGGGGCTCCCGCGAGGGGCGAGAGAGAG 1025  
 QY 141 TTPGlyGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGln 160  
 DB 1026 TGGGAGACGACACAG 1085  
 QY 161 LysAlaGlnHsTyrLysProTyrTyrLysLeuThrTTPGlnGlnGlnGlnGlnGlnGln 180  
 DB 1086 AAGCGGCATTGGAAACCGTACTCAAGCTCGAGCTGGAGAGAGAGAGAGAGAGAGAG 1145  
 QY 181 LysGlnSerLeuArgAlaSerArgLysArgAlaGlnMetPheAlaLysGlnGlnProVal 200  
 DB 1146 AAACAGAGCTTCGAGACTTCAGAGATCCAGAGCATATTTGACCAAGAGGCCAGCGGATC 1205  
 QY 201 AlaProTyrAsnThrThrGlnPheLeuMetCAspPheAspGlnGlnGlnGlnGlnGln 220  
 DB 1206 GCGCCTTAACACACAG 1265  
 QY 221 LysThrGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGln 240  
 DB 1266 AAACCGGCTGTACTCCAAAGCGCGCGCGCCCAATCCGACGACACCGAGATGACGAC 1325  
 QY 241 PheMetGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGln 260  
 DB 1326 TTCATGAGAAAGAGGGGTGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1385  
 QY 261 GluPheLeuGlnArgAspPheSerGlnThrTyrGlnGlnGlnGlnGlnGlnGlnGlnGln 280  
 DB 1386 GAGTTTCTCCAGCGGAGCTTCTCGAGAGCTGACAGCGGTACACACGAGAGAGCTCGCAG 1445  
 QY 281 AsnMetSerLysGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGln 300  
 DB 1446 AACATGAGCAACAGAGAGCTCATCAAGAGGTACTGGAAGCTGGAAGAGTGGCTCTGCGC 1505  
 QY 301 MetGlnAspGlnAsnAsnArgLeuArgLeuGlnSerLysArgLeuGlnGlnGlnGlnGln 320  
 DB 1506 ATGAGAGAGAGAAACACCGGCTGCGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1565  
 QY 321 ArgValArgGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGln 340  
 DB 1566 CCGTGCAG 1625  
 QY 341 ThrGlnAsnGlnLeuHsArgGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGln 359

Db	1626	ACCGAGAACGACTGCACC	GCGCAGCAGGACGGCCGCTTTTCCAAGTTTGAGAC	1682
RESULT 2				
LOCUS	BD106410		2199 bp	DNA linear PAT 18-SEP-2002
DEFINITION	Secreted proteins and polynucleotides encoding them.			
ACCESSION	BD106410			
VERSION	BD106410.1		GI:23201228	
KEYWORDS	JP 2002503955-A/1.			
SOURCE	Chlamydia sp.			
ORGANISM	Chlamydia sp.			
REFERENCE	Bacteria; Chlamydiae; Chlamydiales; Chlamydiaceae; Chlamydia.			
AUTHORS	1 (bases 1 to 2199)			
TITLE	Jacobz,K., McCoy,J.M., Lavallie,E.R., Racie,L.A., Merberg,D., Treacy,M., Spaulding,V. and Agostino,M.J.			
JOURNAL	Secreted proteins and polynucleotides encoding them			
COMMENT	Parent: JP 2002503955-A 1 05-FEB-2002;			
	GENETICS INSTITUTE INC			
	PJ	JP 2002503955-A/1		
	PD	05-FEB-2002		
	PF	20-MAR-1998	JP 1998545874	
	PR	21-MAR-1997	US 08/822167, 19-MAR-1998	US 09/044466
	KENNETH JACOBS, JOHN M MCCOY, EDWARD R LAVALLIE, LISA A RACIE, PI DAVID MERBERG,			
	PI MAURICE TREACY, VIKKI SPAULDING, MICHAEL J AGOSTINO PC			
	C12N15/12,C07K14/47,A61K38/17			
	CC Strandedness: Double;			
	CC Topology: Linear;			
FEATURES				
source	Location/Qualifiers.			
	1..2199			
	/organism="Chlamydia sp."			
	/mol_type="genomic DNA"			
	/db_xref="taxon:35827"			
BASE COUNT	552 a	511 c	674 g	462 t
ORIGIN				
Alignment Scores:				
Pred. NO.:	1.34e-92	Length:	2199	
Score:	1910.00	Matches:	359	
Percent Similarity:	100.00%	Conservative:	0	
Best Local Similarity:	100.00%	Mismatches:	0	
Query Match:	100.00%	Indels:	0	
DB:	6	Gaps:	0	
US-09-972-758A-2 (1-359) x BD106410 (1-2199)				
QY	1	MetalagluPProphelenuSerGlutrycIinhisglnPProginThrSerAsnCysThrily	20	
Db	707	ATGCCCGAGCATTTCTGTTCAGAATATCAACACCGCTTCAAACTAGCAACTGTTCACAGT	766	
QY	21	AlalaalaValGLinglucIuleuasnProgluaAgPproProglialaGLuglunargIa	40	
Db	767	GCTGCTGCTGTCACAGAAAGCTGAACCTGAGCCCCCAAGCGCGAGAGCGGGGTG	826	
QY	41	ProglugluaspSerArgTrpGlnSerArglaPheProginLeuGLyIYAqProgly	60	
Db	827	CCCAGAGGACACAGTAgGTGCAATCGAAGAGGTTCCCCCAATTGGTGGCCGTCCGGGG	886	
QY	61	ProgluglygluylSerleuGluSerGlnProProProleuGlnInThrglnaALCysePro	80	
Db	887	CCGGAGGGGGAAGGAGGCTTGAAATCCCAACCACTCCCTTGACAGACCAGGCTGTCCA	946	
QY	81	GluseSerCybleuarArgGluglulyleGlyglinaEngIYAapAspSerSerlaGly	100	
Db	947	GAATTACCTGCTCTAGAGAGGCGCAGAAAGGGCCGAATAGGGGAGACTCGTCCGCTGGC	1006	
QY	101	GIYAapPheProProProAlaGluvalaGluProThrProglualaGluLeuLeualaGln	120	
Db	1007	GCGCACTTCCCGCGCCGCGCAGAAATGGAAACCGACGCCCGAGGCGCACTGCTGCCACG	1066	
QY	121	ProCysHIAapSPserGluLaLaserlysleuGlylaProilaalaglyGlygluGlu	140	

Dn		1067	CCTTGTGATACCTCCAGGCCAAGTAAGTTGGGGGCTCCTCCGAGGGGGCGAAGAAGAG	1126
Oy		141	TTPGLYGLNGINGLINAARGINEUGLYLYVALYSHISARGARG-PROSERLYLS	160
Dn		1127	TGGGAGCAGCAGCAGADACAGCTGGGGAAAGAAAAACATGAGAGACCCTGCMAAG	1186
Oy		161	LYEARGHSITRPLYSERPLYTYTRYVLEUTHRTPLGUGLUWLYSBLVSYPHEASPGLU	180
Dn		1187	AAGGGGCAATTGGAAAACGGTACTCACMACCTACCTGGGAAGAAAGAAAGTTTCAGCAG	1246
Oy		181	LYEGLINSEIRUEURGALASERARGILEARXALAGLUNEC PHEALAYSGLYGINPRVAI	200
Dn		1247	AAACAGAGCCTTGAGCTTCAAAGGATCCGAGCCGAGATGTTCCGCAAGGGCCAGCCGTC	1306
Oy		201	ALAProTYRANThrThrGlnPheLeuMetAspAsphHisaspGlnGluProaspLeu	220
Dn		1307	GCGCCCTATAACAACACGACGAGTTCTCATGATGATCACGACCGAGAGCCGATCTC	1366
Oy		221	LYSTHRLYLTYRYSERYLRARGALALALALYSESRNAPHTHRSETRASPASP	240
Dn		1367	AAAACCGGCTGTACTCCAAGCGGGCCGCCAATCCAGCAGCACACAGATGACGAC	1426
Oy		241	PHEMEGLUGLUGLYGIUGLUAASPGILGYISERASPGLYMETGYLYIASPGIYser	260
Dn		1427	TTCAATGAAAGAGGGGTGAGAGATGGGGGACCATGGATGGAGGGGACGGCAGC	1486
Oy		261	GLUPHEUGLINARGASPPNESERIUTHRTRYGLUARQTYRHISTHRLGUSERLEUGLN	280
Dn		1487	GAGTTTCTGCAGCGGGACTTCTCGGAGACGATACGAGCGGATACACACGAGAGCCTCAG	1546
Oy		281	ASNMECSERYSGNLULEULLELYGLUTRYLEUGLULEUGLIULYCYLEUSERFARG	300
Dn		1547	AACATGAGCAAGCAGGAGCTCATCAAGAGATCACTGGAACTGGAGAAATGCTCTCCGC	1606
Oy		301	METGLUASPGLUASANAARGLEUARGLEUGLUSERYRSARGLEUGLYLYIASPASPALA	320
Dn		1607	ATGAGGAGACAGAACCAACCGGCTGGCTGGAGAGCAAGCGGCTGGGTGGCGACGACCG	1666
Oy		321	ARYGVALRGSLUGLUGLULEUGLULEUASPARGLEUARGLAAGLUAENLEUGLNULEU	340
Dn		1667	CGTGTGGGGAGAGCTGGAGCTGGAGTGGACCGGCTGGCGCGCGAAMCCTTCACTCCTG	1726
Oy		341	THRTGLUANGLUENHISARGLINGLINGLUARGALAPROLEUSERLYSPHEGLYASP	359
Dn		1727	ACCAGAACCAACTGCACCGGACGACGAGCGAGCGCGCTTCCAAATTGGAGAC	1783
RESULT 3				
AB021179				
LOCUS	AB021179	Homo sapiens mRNA for HEXIM1 protein, complete cds.	3624 bp	mRNA linear PRI 05-DEC-2000
DEFINITION	AB021179			
ACCESSION	AB021179.1	GI:4062855		
VERSION	HEXIM1; HEXIM1 protein.			
KEYWORDS		HOMO SAPIENS		
SOURCE		Homo sapiens (human)		
ORGANISM		Homo sapiens		
REFERENCE		Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;		
AUTHORS		Mammalia; Eutheria; Primates; Catarrhini; Homidae; Homo.		
TITLE		1 (bases)		
JOURNAL		Kusuhara, M., Nagasaki, K., Kimura, K., Maase, N., Manabe, T.,		
REFERENCE		Ishikawa, S., Akawa, M., Miyazaki, K. and Yamaguchi, K.		
TITLE		Cloning of hexamethylene-bis-acetamide-inducible transcript,		
AUTHORS		HEXIM1, in human vascular smooth muscle cells		
JOURNAL		Biomed. Res. 20, 273-279 (1999)		
REFERENCE		2 (bases 1 to 3624)		
TITLE		Kusuhara, M.		
AUTHORS		Direct Submission		
JOURNAL		Submitted (14-DEC-1998) Masatoshi Kusuhara, National Defense		
REFERENCE		Medical College, The First Department of Internal Medicine; 3-2		
TITLE		Namik, Tokorozawa, Saitama 359-8513, Japan		
AUTHORS		(E-mail:mku@med.nidmc.ac.jp, Tel:+81-42-995-1597,		
REFERENCE		Fax:+81-42-996-5200)		
TITLE		Location/Qualifiers		

```

source
1. .3624
/organism="Homo sapiens"
/mol_type="mRNA"
/db_xref="taxon:9606"
/cell_type="coronary artery smooth muscle cell"
1. .3624
/gene="HEXIM1"
690. .1769
/codon_start=1
/product="HEXIM1 protein"
/protein_id="BAA36166.1"
/db_xref="GI:4062856"
/translation="MAEPLISEYOHQPTNCTGAAGVQEBINPERPGAEERVPEED
SRWGRAPFQLEGPGESESLSPPLQTOACPESECLREKQKGGHRSRSEK
PPPAVEPTPAELLAPCHDSEASKLAPAPAGEEENQOOROLGKKHRRSRSEK
KHKMPYVLTWEKKKPEKOSLPAKRAMPAKGGVAPYNTTPIFMDHDEEP
DKTGLYSKRAAKSDTSDDDPMEBEGSEDSGDMGDSGEFLORDPSEYERHT
ESLQNNKQELIKEYLEKCSRMEDENNRRLSEKRLGSDARVRELELDRLRA
ENIQLTENELHROQERAPLSKFGD"

BASE COUNT      978 a      775 c      964 g      907 t
ORIGIN

Alignment Scores:
Pred. No.:      2, 22e-92      Length:      3624
Score:          1910.00      Matches:      359
Percent Similarity: 100.00%      Conservative: 0
Best Local Similarity: 100.00%      Mismatches: 0
Query Match:    100.00%      Indels:      0
DB:              Gaps:      0

US-09-972-758a-2 (1-359) x AB021179 (1-3624)

QY      1 MetAAGLPurPheUeSerGIuTYrGInhiAGInProGInthSerAenCyThrGIy 20
DB      690 ATGGCCGAGCCATTCTTGTCAGATATCAACACAGCCCTCAACTGACAACTGTACAGG 749
QY      21 ALaAaLaVaIGInGIuLeuAnProGluArProProGIyAAGluGIuArGva 40
DB      750 GCTGCTGCTGTCAGAGAAAGAGCTGAACTGAGCGGCCCGCAGCGCGAGAGAGCGGG 809
QY      41 ProGluGIuAerPheSerArGIrPInSerArGAlAaPheProGInleuGIyGIyArProGIy 60
DB      810 CCCGAGGAGAGACAGTGTGGTGCATAGAGAGCGCTTCCCAAGTGGGTGGCCCTCCGGGG 869
QY      61 ProGluGIuGIySerleuGIuSerGIuProProProleuGIuInthGIuAaCySPro 80
DB      870 CCGAGAGGGAGAGAGCGCTGGAATCCCAACCTCCCTTGCAGACCCAGGCTGTCCA 929
QY      81 GluSerSerCySeuArGIuGIyGIuGIyGIuGIyAAsPaerSerArGIy 100
DB      930 GAATCTAGCTGCTGAGAGAGAGCGCGAGAGCGCGCAAAATGGGAGCAACTGCTCGCTGGC 989
QY      101 GIyAerPheProProProAAGluVaIGluProthProGluAAGluLeuAaGIu 120
DB      990 GCGGACTTCCCGCGCGCGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1049
QY      121 ProCyHiaAerPheSerGIuAerGIySerleuGIyAaProAaAAGluGIyGIuGIu 140
DB      1050 CTTGTCAAGACTCCAGAGCAAGTAAGTGGGGCTCTGCGCGAGAGAGAGAGAGAGAG 1109
QY      141 TTPGIGInGIuGIuGIuGIuGIuGIyGIyGIyGIyGIyGIyGIyGIyGIyGIyGIy 160
DB      1110 TGGGAGACAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1169
QY      161 LySArGIhETrLyAerProTYrTYrLySleuThrTTPGIGInGIuGIyGIyGIyGIy 180
DB      1170 AAGCGGCAATTGGAAACCGTACTCAAGAGCTGAGTGGAGAGAGAGAGAGAGAGAGAG 1229
QY      181 LyGIInSerleuAaAaSerArGIleAGlAAGluMeCPheAlAaLySglYGIuInProAl 200
DB      1230 AAACAGAGGCTTCAGACTTCAGAGATCCAGAGGATGTTCGCAAGAGAGAGAGAGAG 1289

```

```

QY      201 ALaProTYrAerThrThrGIuPheUeNeCtAaPaerPhIAerGIuGIuGIuProAaPleu 220
DB      1290 GCGCCCTATTAACACAGAGAGAGTCTCTCAAGTATGATATCAAGAGAGAGAGAGAGAG 1349
QY      221 LySThrGIyLeuTYrSerleuAaGIuAaAaAaAaAaAaAaAaAaAaAaAaAaAaAa 240
DB      1350 AAACCGGCTCTACTCCAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1409
QY      241 PheMeGIuGIuGIyGIyGIuGIuAaPgiGIySerAaPgiYMeGIyGIyAaPgiYSer 260
DB      1410 TTCAATGAGAAAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1469
QY      261 GIuPheUeGIuAaAaPheSerGIuThrTYrGIuAaGIyGIyGIyGIyGIyGIyGIy 280
DB      1470 GAGTTTCTGACAGCGGAGACTTCTCGAGAGCTGACAGCGGTACACAGAGAGAGAGAG 1529
QY      281 AaNeCtSerleuGIuGIuGIuGIuGIuGIyGIyGIyGIyGIyGIyGIyGIyGIyGIy 300
DB      1530 AACATAGAGACAGAGAGAGCTCATCAAGAGATACCTGGAAGTGAAGAGAGAGAGAGAG 1589
QY      301 MeCtGIuAerGIuAaAaAaAaAaAaAaAaAaAaAaAaAaAaAaAaAaAaAaAaAaAa 320
DB      1590 ATGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1649
QY      321 ArGvaAaAaGIuLeuGIuGIuGIuGIuAaAaAaAaAaAaAaAaAaAaAaAaAaAaAa 340
DB      1650 CGTGTGCGGAGAGCTGAGAGCTGAGAGCTGAGAGCTGAGAGCTGAGAGAGAGAGAGAG 1709
QY      341 ThrGIuAaenGIuLeuHiaerGIuGIuGIuGIuGIuGIuGIuGIuGIuGIuGIuGIuGI 359
DB      1710 ACCGAGAAAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1766

RESULT 4
AC142472/c 35050 bp DNA linear PRI 04-APR-2003
LOCUS
DEFINITION
AC142472
ACCESSION
AC142472.1 GI:29540490
VERSION
KEYWORDS
SOURCE
ORGANISM
Homo sapiens (human)
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
1 (bases 1 to 35050)
REFERENCE
Birren, B., Nuebaum, C. and Lander, B.
Homo sapiens chromosome 17, clone XH105-811963
Unpublished
2 (bases 1 to 35050)
Birren, B., Nuebaum, C., Lander, B., Aboueleil, A., Allen, N.,
Anderson, S., Arachchi, H. M., Barna, N., Bastien, V., Bloom, T.,
Boguslavsky, L., Boukhalter, B., Camarata, J., Chang, J., Choepel, Y.,
Collamore, A., Cook, A., Cooke, P., Cornu, B., Deavello, J.,
Diaz, J. S., Dodge, S., Dooley, K., Dorris, L., Erickson, J., Faro, S.,
Ferreira, P., Fitzgerald, M., Gage, D., Galagan, J., Gerdy, S.,
Graham, L., Grand-Pierre, N., Hafez, N., Haploplan, D., Hago, B.,
Hall, J., Horton, L., Hulme, W., Iliev, I., Johnson, R., Jones, C.,
Kamat, A., Karatas, A., Kells, C., Landers, T., Levine, R.,
Lindblad-Toh, K., Liu, G., Lui, A., Mabbitt, R., McLellan, C.,
Macdonald, P., Major, J., Manning, J., Matthews, C., McCarthy, M.,
Meltilin, J., Meneses, L., Mihova, T., Mlenga, V., Murphy, T., Naylor, J.,
Nguyen, C., Nicol, R., Norbu, C., O'Connor, T., O'Donnell, P.,
O'Neill, D., Oliver, J., Peterson, K., Phunkhang, P., Pierre, N.,
Rachupka, A., Ramsamy, U., Raymond, C., Retta, R., Rise, C., Rogov, P.,
Roman, J., Schauer, S., Schuback, R., Seaman, S., Severy, P., Smith, C.,
Spencer, B., Stange-Thomann, N., Stojanovic, N., Stubbs, M.,
Talamas, J., Teste, S., Theodore, J., Topham, K., Travers, M.,
Vassiliev, H., Venkataraman, V. S., Viet, R., Vo, A., Wilson, B., Wu, X.,
Wymann, D., Young, G., Zainoun, J., Zembek, L., Zimmer, A. and Zody, M.
Direct Submission
TITLE
Submitted (04-APR-2003) Whitehead Institute/MIT Center for Genome
Research, 320 Charles Street, Cambridge, MA 02141, USA
All repeats were identified using RepeatMasker:
Smit, A.F.A. & Green, P. (1996-1997)
COMMENT

```

```

http://ftp.genome.washington.edu/RM/RepeatMasker.html
----- Genome Center
Center: Whitehead Institute/ MIT Center for Genome Research
Center code: MIBR
Web site: http://www-seq.wi.mit.edu
Contact: sequence_submissions@genome.wi.mit.edu
----- Project Information
Center Project name: L29477
Center clone name: 8119_G_3
-----

```

Only the last 35.1 kilobases of this clone are being submitted.  
The remainder overlaps accession number AC008105 [WICGR project  
L890].

## FEATURES

```

source          Location/Qualifiers
1..35050
/organism="Homo sapiens"
/mol_type="genomic DNA"
/db_xref="taxon:9606"
/chromosome="17"
/map="17"
/clone="XXfos-8119G3"
/clone_lib="Human Female Fosmid Library"
14..320
/rpt_family="AluSp"
repeat_region
1947..1983
/rpt_family="AT_rich"
repeat_region
3469..3493
/rpt_family="(A)n"
repeat_region
3993..4099
/rpt_family="C-rich"
repeat_region
5028..5069
/rpt_family="G-rich"
repeat_region
complement(5605..5818)
/rpt_family="MIR"
repeat_region
complement(6064..6196)
/rpt_family="L2"
repeat_region
complement(6713..6862)
/rpt_family="MIR"
repeat_region
6863..7166
/rpt_family="AluSc"
repeat_region
complement(7167..7225)
/rpt_family="MIR"
8048..8104
/rpt_family="(CCCCG)n"
repeat_region
8113..8149
/rpt_family="GC-rich"
repeat_region
9472..9732
/rpt_family="L1MC4"
repeat_region
9733..10036
/rpt_family="AluSg1"
repeat_region
10037..10091
/rpt_family="L1MC4"
repeat_region
complement(10169..10477)
/rpt_family="AluSx"
10480..10739
/rpt_family="L1MC4"
repeat_region
10750..10906
/rpt_family="L1MC5"
repeat_region
complement(10907..11182)
/rpt_family="AluJo"
11183..11552
/rpt_family="L1MC5"
repeat_region
complement(11556..11857)
/rpt_family="AluSx"
repeat_region
complement(11913..12044)
/rpt_family="FLAM_C"
12108..12240
/rpt_family="FLAM_C"
repeat_region
12296..12374
/rpt_family="FLAM_C"
repeat_region
12353..12373
/rpt_family="AluS"
repeat_region
/note="<30 qual SNGL region"
unare

```

```

12374
/note="unresolved sequence region:
100 bp of sequence missing here"
repeat_region
12375..12673
/rpt_family="AluSc"
repeat_region
13055..13232
/rpt_family="MIR"
repeat_region
complement(14041..14336)
/rpt_family="AluSx"
repeat_region
complement(14518..14598)
/rpt_family="MIR"
17580..17884
/rpt_family="AluSx"
repeat_region
17893..18028
/rpt_family="L1MA3"
repeat_region
complement(18032..18350)
/rpt_family="AluY"
18360..18383
/rpt_family="AT-rich"
repeat_region
complement(18387..18515)
/rpt_family="FLAM_C"
18537..18707
/rpt_family="FRAM"
repeat_region
18708..18877
/rpt_family="L1MA4"
repeat_region
19015..19324
/rpt_family="AluVa5"
repeat_region
20221..20288
/rpt_family="(CCCCG)n"
20403..20429
/rpt_family="GC-rich"
repeat_region
complement(21264..21365)
/rpt_family="MIR"
repeat_region
complement(21662..21734)
/rpt_family="L2"
repeat_region
complement(22401..22550)
/rpt_family="L2"
repeat_region
complement(23139..23338)
/rpt_family="MIR"
repeat_region
complement(23346..23673)
/rpt_family="AluJo"
repeat_region
complement(23704..23903)
/rpt_family="MIR"
repeat_region
complement(23996..24125)
/rpt_family="MER53"
24057..24138
/rpt_family="MER53"
repeat_region
25014..25095
/rpt_family="(CA)n"
25274..25344
/rpt_family="MIR3"
repeat_region
complement(25687..25742)
/rpt_family="L2"
repeat_region
complement(25847..25917)
/rpt_family="L1MB8"
25918..26185
/rpt_family="AluJo"
repeat_region
complement(26186..26209)
/rpt_family="L1MB8"
26243..26497
/rpt_family="AluYc"
repeat_region
26500..26799
/rpt_family="AluY"
repeat_region
complement(26800..27209)
/rpt_family="L1MB4"
repeat_region
complement(27220..27528)
/rpt_family="AluSx"
repeat_region
complement(27533..27592)
/rpt_family="AluSg/x"
repeat_region
complement(27597..27800)
/rpt_family="AluSx"
repeat_region
complement(27805..27916)
/rpt_family="AluSg/x"

```

```

repeat_region complement (27917..27963)
repeat_region /rpt_family="LMB4"
repeat_region 29143..29261
repeat_region /rpt_family="MERSB"
repeat_region complement (29547..29628)
repeat_region /rpt_family="L2"
repeat_region complement (30172..30304)
repeat_region /rpt_family="MIR"
repeat_region 30327..30450
repeat_region /rpt_family="MIR"
repeat_region complement (30485..30611)
repeat_region /rpt_family="Alusg/x"
repeat_region 30675..30825
repeat_region /rpt_family="L3b"
repeat_region complement (30871..31136)

Alignment Scores:
Pred. No.: 2,16e-91 Length: 35050
Score: 1910.00 Matches: 359
Percent Similarity: 100.00% Conservative: 0
Best Local Similarity: 100.00% Mismatches: 0
Query Match: 100.00% Indels: 0
DB: Gaps: 0

US-09-972-758a-2 (1-359) x AC142472 (1-35050)

QY 1 MetAgluPProPheLeuSerGluTyrglnHsGlnProGlnHrSerAenCySthngly 20
Db 3397 ATGGCGAGGCACTTCTTGAGAAATCAACACAGGCTCAAACTGACCACTGACGCT 3338

QY 21 AAlAAlAAlAAlGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGln 40
Db 3337 GCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 3278

QY 41 ProGlnGlnUspSerArgTPrGlnSerArgAlaPheProGlnGlnGlnGlnGlnGlnGln 60
Db 3277 CCGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGG 3218

QY 61 ProGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGln 80
Db 3217 CCGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGG 3158

QY 81 GluSerSerCysLeuArgGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGln 100
Db 3157 GAATCTAGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 3098

QY 101 GlysAspPheProProAlaGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGln 120
Db 101 GlysAspPheProProAlaGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGln 101
3097 GGGGAGCTCCCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 3038

QY 121 ProCysHisAspSerGlnUaSerLysLeuGlyAlaProAlaAlaGlnGlnGlnGlnGln 140
Db 3037 CTTTGTATATCTCCGAGGCCAGTAAGTTGGGGCTCTCCCGCGAGGGCGCGAGAGGAG 2978

QY 141 TrpGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGln 160
Db 2977 TGGGAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGC 2918

QY 161 LysArgHisTrpLysProTyrrTyrrLysLeuThrTrpGlnGlnGlnGlnGlnGlnGln 180
Db 2917 AACCGGAGATTGGAACCGTACTCAAGCTAGCTGGGAGAGAGAGAGAGAGAGAGAGAG 2858

QY 181 LysGlnSerLeuArgAlaSerArgIleArgAlaGlnUwePheAlaLysGlnGlnProAl 200
Db 2857 AAACAGAGCTTTCGAGCTTTCAGAGATCCGAGCGAGATGTTCCGCAAGGGCCGCGGTC 2798

QY 201 AlaProTyrrAnthrThrglnPheLeuMetAspAspHisAspGlnGlnGlnGlnGlnGln 220
Db 2797 GCGCCCTATACACACAGCAGCTTCCTCATGATGATACAGACAGGAGGAGGCGGATCTC 2738

QY 221 LysThrgLysLeuTyrrSerLysArgAlaAlaAlaLysSerAspAspHisAspAspAsp 240
Db 2737 AAAACCGGCTGTTACTCAAGCGGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 2678

```

```

QY 241 PhemeGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGln 260
Db 2677 TTCATGAGAGAGAGGAGTGAAGATGGGAGCAGAGGATGGATGGAGGAGGAGCGGACG 2618

QY 261 GluPheLeuGlnArgAspPheSerGluThrTyrrGlnArgTyrrHisThrLysLeuGln 280
Db 2617 GAGTTTCTCAGCGGAGCTTCTCGAGAGTACGAGCGGTATACACAGGAGAGCTCGCAG 2558

QY 281 AsnMetSerLysGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGln 300
Db 2557 AACATAGCAGACAGAGAGCTCATCAAGAGTACCTGGAAGTGAAGAGAGGCTCTCGCC 2498

QY 301 MetGlnAspGlnAsnAsnArgLeuArgLeuGlnSerLysArgLeuGlnGlnGlnGlnGln 320
Db 2497 ATGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 2438

QY 321 ArgValArgGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGln 340
Db 2437 CCGTGGCGGAGAGCTGAGCTGAGAGCTGAGAGCTGCGCGCGCGAGAACTCCAGCTGCTG 2378

QY 341 ThrGlnAsnGlnLeuHisArgGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGln 359
Db 2377 ACCGAGACCAACTGCACTCCGCGAGCAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 2321

RESULT 5
AC138150 38849 bp DNA linear PRI 14-FEB-2003
LOCUS Homo sapiens chromosome 17, clone RP13-890H12, complete sequence.
DEFINITION AC138150
ACCESSION AC138150.4 GI:28376773
VERSION HTG.
KEYWORDS Homo sapiens (human)
SOURCE Homo sapiens
ORGANISM Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
1 (bases 1 to 38849)
REFERENCE Birren, B., Nussbaum, C., Lander, E., Allen, A., Allen, N., Anderson, S.,
Barna, N., Bastien, V., Bloom, T., Boguslavsky, L., Boukhalter, B.,
Camata, J., Chang, J., Chazaro, B., Choepel, Y., Collamore, A.,
Cook, A., Cooke, P., DeArillano, K., Dewar, K., Diaz, J.S., Dodge, S.,
Faro, S., Ferreira, P., Fitzgerald, M., Gage, D., Galagan, J.,
Gardyna, S., Gord, S., Graham, L., Grand-Pierre, N., Hafez, N.,
Hagos, B., Horton, L., Hulme, W., Iliev, I., Johnson, R., Jones, C.,
Kamat, A., Karatae, A., Kells, C., Landers, T., Levine, R.,
Lindblad-Toh, K., Liu, G., Maclean, C., Macdonald, P., Major, J.,
Matthews, C., McCarthy, M., Meldrum, J., Meneu, L., Mihova, T.,
Norman, C.H., O'Connor, T., O'Donnell, P., O'Neill, D., Oliver, J.,
Peterson, K., Phunhang, P., Pierre, N., Raymond, C., Retta, R.,
Rise, C., Rogov, P., Roman, J., Roy, A., Schauer, S., Schnupack, R.,
Seaman, S., Severy, P., Smith, C., Spencer, B., Stange-Thomann, N.,
Stojanovic, N., Talamas, J., Testafy, S., Theodore, J., Topham, K.,
Travers, M., Vassiliev, H., Viel, R., Vo, A., Wilson, B., Wu, X.,
Wyman, D., Young, G., Zainoun, J., Zembek, L., Zimmer, A. and Zody, M.
Direct Submission
Submitted (18-DEC-2002) Whitehead Institute/MIT Center for Genome
Research, 320 Charles Street, Cambridge, MA 02141, USA
3 (bases 1 to 38849)
REFERENCE Birren, B., Nussbaum, C., Lander, E., Abouneillel, A., Allen, N.,
Anderson, S., Arachchi, H.M., Barna, N., Bastien, V., Bloom, T.,
Boguslavsky, L., Boukhalter, B., Camata, J., Chang, J., Choepel, Y.,
Collamore, A., Cook, A., Cooke, P., Corum, B., DeArillano, K.,
Diaz, J.S., Dodge, S., Dooley, K., Dorris, L., Erickson, J., Faro, S.,
Ferreira, P., Fitzgerald, M., Gage, D., Galagan, J., Gardyna, S.,
Graham, L., Grand-Pierre, N., Hafez, N., Hagopian, D., Hagos, B.,
Hall, J., Horton, L., Hulme, W., Iliev, I., Johnson, R., Jones, C.,
Kamat, A., Karatae, A., Kells, C., Landers, T., Levine, R.,
Lindblad-Toh, K., Liu, G., Lui, A., Mabbitt, R., Maclean, C.,

```



TITLE  
JOURNAL  
AUTHORS  
REFERENCE

Macdonald, P., Major, J., Manning, J., Matthews, C., McCarthy, M.,  
Meldrum, J., Meneus, L., Mihova, T., Mienga, V., Murphy, T., Naylor, J.,  
Nguyen, C., Nicol, R., Norbu, C., O'Connor, T., O'Donnell, P.,  
O'Neill, D., Oliver, J., Peterson, K., Phunkhang, P., Pierre, N.,  
Rachupka, A., Ramasamy, U., Raymond, C., Retta, R., Rise, C., Rogov, P.,  
Roman, J., Schauer, S., Schupbach, R., Seaman, S., Severy, P., Smith, C.,  
Spencer, B., Stange-Thomann, N., Stojanovic, N., Stubbs, M.,  
Talamas, J., Tesfaye, S., Theodore, J., Topham, K., Travers, M.,  
Vassiliev, H., Venkataraman, V.S., Viel, R., Vo, A., Wilson, B., Wu, X.,  
Wyman, D., Young, G., Zainoun, J., Zembek, L., Zimmer, A. and Zody, M.  
Direct Submission  
Submitted (31-0AN-2003) Whitehead Institute/MIT Center for Genome  
Research, 320 Charles Street, Cambridge, MA 02141, USA  
4 (bases 1 to 38849)

Anderson, S., Arachchi, H.M., Barna, N., Bastien, V., Bloom, T.,  
Boguslavskiy, L., Boukhalter, B., Camarata, J., Chang, J., Choepel, Y.,  
Collymore, A., Cook, A., Cooke, P., Corum, B., Dextrilliano, K., Faro, S.,  
Diaz, J.S., Dodge, S., Doolley, K., Dorris, L., Erickson, J., Fero, S.,  
Ferreira, P., FitzGerald, M., Gage, D., Galaan, J., Gardyna, S.,  
Graham, L., Grand-Pierre, N., Hafez, N., Hagopian, D., Hagos, B.,  
Hall, J., Horton, L., Hulme, W., Iliev, I., Johnson, R., Jones, C.,  
Kamat, A., Karatas, A., Kellis, C., Landers, T., Levine, R.,  
Lindblad-Toh, K., Liu, G., Lui, A., Mabbitt, R., Maclean, C.,  
Macdonald, P., Major, J., Manning, J., Matthews, C., McCarthy, M.,  
Meldrum, J., Meneus, L., Mihova, T., Mienga, V., Murphy, T., Naylor, J.,  
Nguyen, C., Nicol, R., Norbu, C., O'Connor, T., O'Donnell, P.,  
O'Neill, D., Oliver, J., Peterson, K., Phunkhang, P., Pierre, N.,  
Rachupka, A., Ramasamy, U., Raymond, C., Retta, R., Rise, C., Rogov, P.,  
Roman, J., Schauer, S., Schupbach, R., Seaman, S., Severy, P., Smith, C.,  
Spencer, B., Stange-Thomann, N., Stojanovic, N., Stubbs, M.,  
Talamas, J., Tesfaye, S., Theodore, J., Topham, K., Travers, M.,  
Vassiliev, H., Venkataraman, V.S., Viel, R., Vo, A., Wilson, B., Wu, X.,  
Wyman, D., Young, G., Zainoun, J., Zembek, L., Zimmer, A. and Zody, M.  
Direct Submission  
Submitted (14-FEB-2003) Whitehead Institute/MIT Center for Genome  
Research, 320 Charles Street, Cambridge, MA 02141, USA  
On Feb 14, 2003 this sequence version replaced gt:28173166.  
All repeats were identified using RepeatMasker:  
Smit, A.F.A. & Green, P. (1996-1997)  
http://ftp.genome.washington.edu/RM/RepeatMasker.html

Genome Center  
Center: Whitehead Institute/ MIT Center for Genome Research  
Center code: WIBR

Web site: http://www-seq.wi.mit.edu  
Contact: sequence\_submissions@genome.wi.mit.edu

Project Information  
Center project name: L28731  
Center clone name: 890\_H\_12

FEATURES  
Source  
Location/Qualifiers  
1..38849  
/organism="Homo sapiens"  
/mol\_type="genomic DNA"  
/db\_xref="taxon:9606"  
/chromosome="17"  
/map="17"  
/clone="RP13-890H12"  
/clone\_id="RP13-890H12 Human Female BAC"  
84..385  
/rpt\_family="AluSx"  
complement(389..758)  
/rpt\_family="L1MC5"  
759..1034  
/rpt\_family="AluDo"  
complement(1035..1191)  
/rpt\_family="L1MC5"  
complement(1202..1461)  
/rpt\_family="L1MC4"  
1464..1772  
/rpt\_family="AluSx"  
complement(1850..1904)  
repeat\_region  
complement(1850..1904)  
/rpt\_family="L1MC4"

repeat\_region  
complement(1905..2208)  
/rpt\_family="AluSg1"  
complement(2209..2469)  
/rpt\_family="L1MC4"  
3792..3828  
/rpt\_family="GC-rich"  
3837..3893  
/rpt\_family="CGGGG)n"  
4716..4774  
/rpt\_family="MIR"  
complement(4775..5078)  
/rpt\_family="AluSc"  
5079..5204  
/rpt\_family="MIR"  
5745..5877  
/rpt\_family="L2"  
6123..6336  
/rpt\_family="MIR3"  
6872..6913  
/rpt\_family="C-rich"  
7842..7948  
/rpt\_family="G-rich"  
8448..8472  
/rpt\_family="(T)n"  
9962..9998  
/rpt\_family="AT-rich"  
complement(11625..11932)  
/rpt\_family="AluSp"  
12103..12389  
/rpt\_family="AluSp"  
complement(12567..12862)  
/rpt\_family="AluSx"  
complement(12863..13169)  
/rpt\_family="AluJb"  
complement(13209..13491)  
13566..13589  
/rpt\_family="AluY"  
13595..13903  
/rpt\_family="AluSp"  
13904..14038  
/rpt\_family="AluSx"  
14039..14352  
/rpt\_family="AluY"  
14353..14522  
/rpt\_family="AluSx"  
14558..14685  
/rpt\_family="FLAM\_C"  
complement(14704..14854)  
/rpt\_family="PRM"  
15205..15307  
/rpt\_family="AluJb"  
15359..15633  
/rpt\_family="AluSx"  
16074..16377  
/rpt\_family="AluSc"  
16380..16558  
/rpt\_family="AluSg/x"  
16743..16877  
/rpt\_family="AluSx"  
16878..17181  
/rpt\_family="AluSp"  
17182..17348  
/rpt\_family="AluSx"  
complement(18100..18399)  
/rpt\_family="AluSc"  
complement(18948..19238)  
/rpt\_family="AluSx"  
22406..22707  
/rpt\_family="AluSx"  
complement(22708..22845)  
/rpt\_family="MIR"  
22960..23228



TITLE  
JOURNAL  
COMMENT

Norbu, C., Norman, C.H., O'Connor, T., O'Donnell, P., O'Neill, D., Oliver, U., Peterson, K., Phunhkhang, P., Pierre, N., Pollara, V., Raymond, C., Retta, R., Rieback, M., Riley, R., Rise, C., Rogov, P., Roman, J., Rosetti, M., Roy, A., Santos, R., Schauer, S., Schuback, R., Seaman, S., Severy, P., Spencer, B., Stange-Thomann, N., Stojanovic, N., Strauss, N., Subramanian, A., Talmas, J., Tafaye, S., Theodore, J., Topham, K., Travers, M., Travis, N., Triggillo, J., Vassiliev, H., Vael, R., Vo, A., Wilson, B., Wu, X., Wyman, D., Ye, W. J., Young, G., Zaitoun, J., Zembek, L., Zimmer, A. and Zody, M.

Submitted (20-APR-2002) Whitehead Institute/MIT Center for Genome Research, 320 Charles Street, Cambridge, MA 02141, USA  
On Apr 20, 2002 this sequence version replaced g1:20177719.  
All repeats were identified using RepeatMasker:  
Smit, A.F.A. & Green, P. (1996-1997)  
<http://ftp.genome.washington.edu/RM/RepeatMasker.html>

Center: Whitehead Institute/ MIT Center for Genome Research

Center code: WIBR

Web site: <http://www-seq.wi.mit.edu>

Contact: [sequence\\_submissions@genome.wi.mit.edu](mailto:sequence_submissions@genome.wi.mit.edu)

Project Information

Center project name: l12020

Center clone name: 403\_G3

Summary Statistics

Sequencing vector: Plasmid; n/a; 100% of reads

Chemistry: Dye-terminator Big Dye; 100% of reads

Assembly program: Phrap; version 0.960731

Consensus quality: 96277 bases at least Q40

Consensus quality: 96051 bases at least Q30

Consensus quality: 98838 bases at least Q20

Insert size: 140000; agarose-fp

Insert size: 99436; sum-of-contigs

Quality coverage: 19.1 in Q20 bases; agarose-fp

Quality coverage: 26.9 in Q20 bases; sum-of-contigs

NOTE: This is a 'working draft' sequence. It currently consists of 15 contigs. Gaps between the contigs are represented as runs of N. The order of the pieces is believed to be correct as given, however the sizes of the gaps between them are based on estimates that have been provided by the submitter.

This sequence will be replaced by the finished sequence as soon as it is available and the accession number will be preserved.

1 8098: contig of 8098 bp in length  
8099 8198: gap of 100 bp  
8199 8271: contig of 73 bp in length  
8272 8371: gap of 100 bp  
8372 9497: contig of 1126 bp in length  
9498 9597: gap of 100 bp  
9598 10863: contig of 1266 bp in length  
10864 10963: gap of 100 bp  
10964 12232: contig of 1269 bp in length  
12233 12332: gap of 100 bp  
12333 13844: contig of 1512 bp in length  
13845 13944: gap of 100 bp  
13945 15215: contig of 1271 bp in length  
15216 15315: gap of 100 bp  
15316 18234: contig of 2919 bp in length  
18235 18334: gap of 100 bp  
18335 21887: contig of 3453 bp in length  
21888 21887: gap of 100 bp  
21888 25485: contig of 3598 bp in length  
25486 25585: gap of 100 bp  
25586 40531: contig of 14946 bp in length  
40532 40631: gap of 100 bp  
40632 52119: contig of 11488 bp in length  
52120 52219: gap of 100 bp  
52220 64941: contig of 12722 bp in length  
64942 65041: gap of 100 bp  
65042 79587: contig of 14546 bp in length  
79588 79687: gap of 100 bp

\* 79688 100836: contig of 21149 bp in length.  
Location/Qualifiers  
source  
1..100836  
/organism="Homo sapiens"  
/mol\_type="genomic DNA"  
/db\_xref="taxon:9606"  
/chromosome="17"  
/map="17"  
/clone.lib="RP11-403G3"  
/clone.lib="RP11-403G3"  
1..8098  
/note="assembly\_fragment"  
clone\_end:SP6  
vector\_side:left"  
8199..8271  
/note="assembly\_fragment"  
8372..9497  
/note="assembly\_fragment"  
9598..10863  
/note="assembly\_fragment"  
10964..12232  
/note="assembly\_fragment"  
12333..13844  
/note="assembly\_fragment"  
13945..15215  
/note="assembly\_fragment"  
15316..18234  
/note="assembly\_fragment"  
18335..21887  
/note="assembly\_fragment"  
21888..25485  
/note="assembly\_fragment"  
25586..40531  
/note="assembly\_fragment"  
40632..52119  
/note="assembly\_fragment"  
52220..64941  
/note="assembly\_fragment"  
65042..79587  
/note="assembly\_fragment"  
79688..100836  
/note="assembly\_fragment"  
BAGE COUNT 23052 a 27082 c 27053 g 22249 t 1400 others  
ORIGIN  
Alignment Scores:  
Pred. No.: 6,24e-91 Length: 100836  
Score: 1910.00 Matches: 359  
Percent Similarity: 100.00% Conservative: 0  
Best Local Similarity: 100.00% Mismatches: 0  
Query Match: 100.00% Indels: 0  
DB: 2 Gaps: 0  
US-09-972-758a-2 (1-359) x AC087298 (1-100836)  
QY 1 MetAaIaGpPheLeuSerGlyYrGlnH:sgInProGlnThrSerAnCySthrgly 20  
DB 70785 ATGGCCGACCAATCTCTTGCAGAAATATCAACACGACCTCAACACTGACAGGT 70726  
QY 21 AAlaAaIaValGInGluGluLeuAnProGluArgProProGlyAlaGluGluArgVal 40  
DB 70725 GCTGCTGCTGCTCAGAGAGAGTGAACCTGAGCCGCCCCAGGGCGGAGAGCGGTG 70666  
QY 41 ProGluGluAaPSeSerArgTrpGlnSerArgAlaPheProGlnLeuGlyGlyYrProGly 60  
DB 70665 CCCGAGGAGGACAGTGAAGTGAATGAGAGCGGTCCCACTTGAGTGGCCGTCCGGGG 70606  
QY 61 ProGluGluGluGlySerLeuGluSerGlnProProLeuGlnThrGlnAlaCySpro 80  
DB 70605 CCGAGAGGGGAGGAGGAGCTGGAATCCCAACCACTCTCTTGCAGACCCAGGCTGTCCA 70546  
QY 81 GluSerSerCyALeuArgGluGluGlyGlyGlyGlnAnGlyAaPSeSerArgAlaGly 100



Db 1316 GCCCCTATTAACACCGAGTTCCTCATGATGATCATCAGACGAGAGCGCGGATCTC 1375  
 QY 221 LyeThrGlyLeuTyrSerLysArgAlaAlaLysSerAspAspThrSerAspAsp 240  
 Db 1376 AAAACCGCGCTGACTTCAAGCGCGCGCGCGCAATCCACACGACGATGACGAC 1435  
 QY 241 PheMetGluGluGlyGlyGluGluAspGlyGlySerAspGlyMetGlyGlyAspGlySer 260  
 Db 1436 TTCATGGAGAGAGGGGTGAGAGGATGGGGGCGACGATGGATGGAGGGAGCGGCGACG 1495  
 QY 261 GluPheLeuGluLysArgAspPheSerGluThrTyrGluLysArgTyrHisThrGluSerLeuGln 280  
 Db 1496 GAGTTTCTGACGGGGGCTTCTCTCGACACCTACGAGCGGTACACACGAGAGCGCTGACG 1555  
 QY 281 AsnMetSerLysGluGluLeuLeuLysGluTyrLeuGluGluGlyGlySerArg 300  
 Db 1556 AACATGACGACGAGGAGCTCATCAGAGTACTGTAACCTGAGAGAGTCTCTCGCGC 1615  
 QY 301 MetGluAspGluLysAsnAsnArgLeuArgLeuGluSerLysArgLeuGlyGlyAspAspAla 320  
 Db 1616 ATGGAGAGACGAGAACCGGCTGGCTGGAGAGCAAGCGGCTGGGTCGACGACGCG 1675  
 QY 321 ArgValArgGluLeuGluLeuGluLeuAspArgLeuArgAlaGluLysLeuGluLeu 340  
 Db 1676 CGTGTGGGAGGCTGGAGCTGGAGCTGGACCGGCTGGCGCCGAGAACCTCCAGCTGCTG 1735  
 QY 341 ThrGluAsnGluLeuHisArgGluGluGluGluArgAlaProLeuSerLysPheGlyAsp 359  
 Db 1736 ACCGAGACCAATCTGCACCGCGACGAGAGCGAGCGCGCTTCCAAATTGGAGAC 1792  
 RESULT 8  
 AK023624 2178 bp mRNA linear PRI 01-AUG-2002  
 LOCUS Homo sapiens cDNA FLJ13562 fis, clone PLACE1008080, highly similar  
 DEFINITION to Homo sapiens mRNA for HEXIM1 protein.  
 ACCESSION AK023624  
 VERSION AK023624.1 GI:10435606  
 KEYWORDS oligo capping; fis (full insert sequence).  
 SOURCE Homo sapiens (human)  
 ORGANISM Homo sapiens  
 Eukaryota; Chordata; Craniata; Vertebrata; Euteleostomi;  
 Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.  
 REFERENCE 1  
 AUTHORS Isegai,T., Ota,T., Hayashi,K., Sugiyama,T., Otsuki,T., Suzuki,Y.,  
 Nishikawa,T., Nagai,K., Sugano,S., Shiratori,A., Sudo,H.,  
 Nagatsuna,M., Hosokita,T., Kaku,T., Kodaira,H., Kondo,H.,  
 Sugawara,M., Takahashi,M., Chiba,Y., Ishida,S., Murakawa,K.,  
 Ono,Y., Takiguchi,S., Watanabe,S., Kimura,K., Murakami,K.,  
 Ishii,S., Kawai,Y., Saito,K., Yamamoto,J., Wakamatsu,A.,  
 Nakamura,Y., Nagahari,K., Masuho,Y., Ninomiya,K. and Iwayanagi,T.  
 NEDO human cDNA sequencing project  
 Unpublished  
 2 (bases 1 to 2178)  
 Isegai,T. and Otsuki,T.  
 Direct Submission  
 Submitted (23-AUG-2000) Takao Isegai, Helix Research Institute,  
 Genomics Laboratory, 1532-3 Yana, Kisarazu, Chiba 252-0812, Japan  
 (E-mail:genomics@hri.co.jp, Tel:81-438-52-3975, Fax:81-438-52-3986)  
 NEDO human cDNA sequencing project supported by Ministry of  
 International Trade and Industry for Japan; cDNA full insert  
 sequencing: Research Association for Biotechnology; cDNA library  
 construction, 5'- & 3'-end one pass sequencing and clone selection:  
 Helix Research Institute (supported by Japan Key Technology Center  
 etc.) and Department of Virology, Institute of Medical Science,  
 University of Tokyo.  
 Location/Qualifiers  
 1..2178  
 /organism="Homo sapiens"  
 /mol\_type="mRNA"  
 /db\_xref="taxon:9606"  
 /clone="PLACE1008080"  
 /rname="PLACE1"  
 /clone\_id="PLACE1"

BASE COUNT 525 a 513 c 678 g 462 t  
 ORIGIN  
 Alignment Scores:  
 Pred. No.: 3,12e-92 Length: 2178  
 Score: 1903.00 Matches: 358  
 Percent Similarity: 99.72% Conservative: 0  
 Best Local Similarity: 99.72% Mismatches: 1  
 Query Match: 99.63% Indels: 0  
 DB: 9 Gaps: 0  
 US-09-972-758a-2 (1-359) x AK023624 (1-2178)  
 QY 1 MetAlaGluProPheLeuSerGluTyrGlnHisGluProGluThrSerAsnCySerGly 20  
 Db 716 ATGGCGGAGCCATTCTTGTCAAGATTCACACACAGCCTCAACTGCAACTGACAGT 775  
 QY 21 AlaAlaAlaValGluGluGluLeuAsnProGluArgProGluValAlaGluGluArgVal 40  
 Db 776 GCTGTGCTGTCCAGAGAGGCTGAACCTGAGCGCCCCCAGCGCGGAGAGCGGGTG 835  
 QY 41 ProGluGluAspSerArgTyrGlnSerArgAlaPheProGluLeuGlyGlyArgProGly 60  
 Db 836 CCGAGAGAGACAGTATGAGTGGCAATGAGAGCGTTCCCACTTGGTGGCGCTCCGGG 895  
 QY 61 ProGluGluGluGlySerLeuGluSerGluGlnProProProLeuGluThrGlnAlaCyPro 80  
 Db 896 CCGAGAGAGAGAGAGGCTTGGATCCAACTCTCTTGCACGCCAGGCTTCTCA 955  
 QY 81 GluSerSerCysLeuArgGluGluGlyGlyGlyGlyGlyGlyGlyGlyGlyGlyGly 100  
 Db 956 GATCTAGCTGCTGCTGAGAGAGGCGGAGAGAGGCGAGATGGGAGCAGACTGCTGCTGCG 1015  
 QY 101 GlyAspPheProProProAlaGluValGluProThrProGluAlaGluLeuLeuAlaGln 120  
 Db 1016 GGGGACTTCCCGCGCGCGGAGAGAGGAGCCAGCCGAGGCTGCTGCTGCTGCG 1075  
 QY 121 ProCysHisAspSerGluLysSerLysLeuGlyAlaProAlaAlaGlyGlyGluGlu 140  
 Db 1076 CTTTGTCAATGATCTCCAGAGCCAGTATGTTGGGGCTCTCTGCGCAGGGGCGAGAGAG 1135  
 QY 141 TrpGluGluGluGluGluGluGluGluGlyGlyGlyGlyGlyGlyGlyGlyGlyGly 160  
 Db 1136 TGGGAGACAGACAGAGAGACAGCTGGGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1195  
 QY 161 LysArgHisTrpLysProTyrTyrLysLeuThrTrpGluGluLysLysPheAspGlu 180  
 Db 1196 AAGCGGCAATTTGAGAAACCTGACTACAACTGACCTGGAGAGAGAGAGAGAGAGAG 1255  
 QY 181 LysGlnSerLeuArgAlaSerArgLysArgAlaGluMetPheAlaLysGlyGlnProVal 200  
 Db 1256 AACAGAGCCTTGCAGACTTCAGATCCAGATCCAGATGTTGCGCAAGGAGCCAGCGGTC 1315  
 QY 201 AlaProTyrAsnThrThrGlnPheLeuMetAspAspHisAspGlnGluProAspLeu 220  
 Db 1316 GCCCCTATTAACACCGAGTTCCTCATGAGATGATCAAGACAGAGAGAGAGCGGATCTC 1375  
 QY 221 LyeThrGlyLeuTyrSerLysArgAlaAlaLysSerAspAspThrSerAspAsp 240  
 Db 1376 AAAACCGCGCTGACTTCAAGCGCGCGCGCGCAATCCACACGACGATGACGAC 1435  
 QY 241 PheMetGluGluGlyGlyGluGluAspGlyGlySerAspGlyMetGlyGlyAspGlySer 260  
 Db 1436 TTCATGGAGAGAGGGGTGAGAGGATGGGGGCGACGATGGATGGAGGGAGCGGCGACG 1495  
 QY 261 GluPheLeuGluLysArgAspPheSerGluThrTyrGluLysArgTyrHisThrGluSerLeuGln 280  
 Db 1496 GAGTTTCTGACGGGGGCTTCTCTCGACACCTACGAGCGGTACACACGAGAGCGCTGACG 1555  
 QY 281 AsnMetSerLysGluGluLeuLysGlyGlyGlyGlyGlyGlyGlyGlyGlyGlyGly 300  
 Db 1556 AACATGACGACGAGGAGCTCATCAGAGTACTGTAACCTGAGAGAGTCTCTCGCGC 1615

Accession	Version	Source	Organism	Reference	Authors	Title	Journal
301	1616	ATGAGAGCGAGAACACCCGGCTGGCTGAGACCAAGCCGCTGGCTGGAGACACCCG	16175				
321	1676	CGTGGCGGAGCTGGAGCTGGAGCTGGAGCTGGAGCTGGAGCTGGAGCTGGAGCTGG	17359				
341	1736	ACCGAAGACGAACTGACCGGAGCGAGCGAGCGAGCGAGCGAGCGAGCGAGCGAGC	1792				
AC024047/c							
AC024047							
AC024047							
AC024047.3	GI:8570008	HTG: HTGS_PHASE1; HTGS_DRAFT.					
Homo sapiens (human)							
Homo sapiens							
Eumalacoidea; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;							
Mammalia; Eutheria; Primates; Catarrhini; Hominiidae; Homo.							
1 (bases 1 to 142326)							
Waterston,R.H.							
The sequence of Homo sapiens clone							
Unpublished							
2 (bases 1 to 142326)							
Waterston,R.H.							
Direct Submission							
Submitted (20-FEB-2000) Genome Sequencing Center, Washington							
University School of Medicine, 444 Forest Park Parkway, St. Louis,							
MO 63108, USA							
On Jun 16, 2000 this sequence version replaced gi:7109599.							
COMMENT							
----- Genome Center -----							
Center: Washington University Genome Sequencing Center							
Center code: MUGSC							
Web site: http://genome.wustl.edu/gsc/index.shtml							
----- Project Information -----							
Center project name: H_NH040303							
----- Summary Statistics -----							
Sequencing vector: Plasmid, 100%							
Sequencing vector: Plasmid, 100%							
Chemistry: Dye-terminator Big Dye 3 of reads							
Chemistry: Dye-terminator Big Dye 3 of reads							
Assembly program: Phrap; version 0.990319							
Consensus quality: 127942 bases at least Q40							
Consensus quality: 13308 bases at least Q30							
Consensus quality: 135734 bases at least Q20							
Insert size: 147000; agarose-fp							
Insert size: 139626; sum-of-contigs							
Quality coverage: 4.06 in Q20 bases; sum-of-contigs							
Quality coverage: 4.33 in Q20 bases; sum-of-contigs							
----- NOTE: This is a 'working draft' sequence. It currently							
* consists of 28 contigs. The true order of the pieces							
* is not known and their order in this sequence record is							
* arbitrary. Gaps between the contigs are represented as							
* runs of N, but the exact sizes of the gaps are unknown.							
* This record will be updated with the finished sequence							
* as soon as it is available and the accession number will							
* be preserved.							
1 1575: contig of 1575 bp in length							
1576 1675: gap of unknown length							
1676 3114: contig of 1439 bp in length							
3115 3214: gap of unknown length							
3215 5389: contig of 2175 bp in length							
5390 5489: gap of unknown length							
5490 7129: contig of 1640 bp in length							
7130 7229: gap of unknown length							

FEATURES	source	Location/Qualifiers
BASE COUNT	32503 a 37098 c 37256 g 32736 t 2733 others	
ORIGIN	/clone="RP11-403G3"	
	/db_xref="taxon:9606"	
	/chromosome="17"	
	/organism="Homo sapiens"	
	/mol_type="genomic DNA"	
ALIGNMENT SCORES:		
Pred. No.:	4,66e-88	Length: 142326
Score:	1858.50	Matches: 357
Percent Similarity:	99.44%	Conservative: 0
Best Local Similarity:	99.44%	Mismatches: 2
Query Match:	97.30%	Indels: 2
DB:	2	Gaps: 0
US-09-972-758A-2 (1-359) x AC024047 (1-142326)		
QY	1	MetaAgglutininProphageLenseGluTyrGlnHisGlnProGlnThrSerAncysThrgly 20
Db	108505	ATGCGCCGAGCATCTTGTTCAGAAATATCAACACGCGCTCAAACTAGCAACTGTACAGGT 108446
QY	21	AlaAlaAlaValGlnGlnGlnLeuAsnProGluAaGpProProGlyAlaGlnGlnAaVal 40

DB 108445 GCTGCTGCTGTCAGAGAGAGCTGAACTCTGAGCCGCCCAAGCGCGAGAGAGCGGGTG 108386  
 QY 41 ProGlulGluAspSerArgTrpGlnSerArgAlaPheProGlnLeuGlyValArgProGly 60  
 DB 108385 CCGGAGAGAGACAGTGTGGCAATGAGAGCGTTCCCAAGTGGTGCGCTCCGGGG 108326  
 QY 61 ProGlulGluGlySerLeuGluSerGlnProProProLeuGlnThrGlnAlaCysPro 80  
 DB 108325 CCGGAGGGGAGAGAGCGCTGGAATCCCAACCACTCCCTTGAGAGCCAGGCTCTCCA 108266  
 QY 81 GluSerSerCysLeuArgGlnGlyGlyValGlnAsnGlyAspAspSerSerAlaGly 100  
 DB 108265 GAATCTAGCTGCTCTGAGAGAGGGCGAGAAAGGGCCAGAAATGGGAGCACTGCTCCGTGGC 108206  
 QY 101 GlyAspPheProProProAlaGluValGluProThrProGlnAlaLeuLeuAlaGln 120  
 DB 108205 GC-GACTTCCTCCGGG-CGGAGAGAAAGTGAACCCAGCCCGAGGCGCAAGCTGCTGCCAG 108148  
 QY 121 ProCysHisAspSerGlnAlaSerLysLeuGlyAlaProAlaAlaGlyGlyGlnGlu 140  
 DB 108147 CCTTGCTGATGACTCCAGAGCGCACTAAGTTGGGGCTCTCCCGCAGGGGCGAGAGAG 108088  
 QY 141 TrpGlyGlnGlnGlnArgGlnLeuGlyLysLysLysHisArgArgArgProSerLysLys 160  
 DB 108087 TGGGACACAGACAGACAGACAGCTGGGGAGAAAGAAAAAATAGAGAGACGCCCTCCAGAAAG 108028  
 QY 161 LysArgHisTrpLysProTyrTyrLysLeuThrTrpGlnGluLysLysLysPheAspGlu 180  
 DB 108027 AACCGGCACTTGGAAACCGTACTCAAGCTGACCTGGAGAAAGAAAAAGTTTCAGAGAG 107968  
 QY 181 LysGlnSerLeuArgLysArgLysArgLysArgLysArgLysArgLysArgLysArgLys 200  
 DB 107967 AAACAGAGCTTGAAGCTTCAAGATCCAGCCGAGATGTTCCCAAGGCGCCAGCGGTC 107908  
 QY 201 AlaProTyrAsnThrThrGlnPheLeuMetAspAspHisAspGlnGluProAspLeu 220  
 DB 107907 GCGCCCTATACCAACGCGAGTTCCATGATGATCAACAGCAGAGAGAGCGGATCTC 107848  
 QY 221 LysThrGlyLeuTyrSerLysArgAlaAlaAlaLysSerAspAspThrSerAspAspAsp 240  
 DB 107847 AAAACCGGCTTGAATCTCCAAAGCGGCGCGCCCAATCCGACACACACAGCTGACGAG 107788  
 QY 241 PheMetGlnGluGlyGlyGlnGluAspGlyGlySerAspGlyMetGlyGlyAspGlySer 260  
 DB 107787 TTATGAGAAAGAGGGGTGAGAGATGGGGGAGCGATGGATGGAGGGGAGCGGAGC 107728  
 QY 261 GluPheLeuGlnArgAspPheSerGluThrTyrGluValGlyTyrHisThrGlnSerLeuGln 280  
 DB 107727 GAGTTTCTGACAGGGGAGCTTCTCGAGAGAGTACGAGGGGTACCAACGAGAGAGCTGACG 107668  
 QY 281 AsnMetSerLysGlnGluLeuLysLysGluTyrLeuGlnLeuGlnLysCysLeuSerArg 300  
 DB 107667 AACATGAGCAGAGAGAGAGCTCATCAAGAGTACCTGGAACCTGGAAGTCTCTCGCGC 107608  
 QY 301 MetGlnAspGluAsnAsnArgLeuArgLeuGlnSerLysArgLeuGlyGlyAspAspAla 320  
 DB 107607 ATGAGAGACAGAAACACCGGCTGGCTGAGAGAGAGAGCGGCTGGTGGCGAGAGCGG 107548  
 QY 321 ArgValArgGlnLeuGlnLeuGlnLeuAspArgLeuArgAlaGluAsnLeuGlnLeu 340  
 DB 107547 CGGTGGCGGAGCTGAGAGCTGAGAGCGGCTGGCGCGCGAGAACTCCAGCTGCTG 107488  
 QY 341 ThrGluAsnGluLeuHisArgGlnGlnGluArgAlaProLeuSerLysPheGlyAsp 359  
 DB 107487 ACCGAGAACCACTGACACCGGCGAGAGAGAGCGCGCTTCCAAATTGGAGAC 107431  
 RESULT 10  
 LOCUS AY090614 3402 bp mRNA linear ROD 18-JUL-2002  
 DEFINITION Mus musculus cardiac lineage protein 1 (Clp1) mRNA, complete cds.  
 ACCESSION AY090614  
 VERSION AY090614.1 GI:20135640  
 KEYWORDS

SOURCE Mus musculus (house mouse)  
 ORGANISM Mus musculus  
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.  
 REFERENCE 1 (bases 1 to 3402)  
 AUTHORS Huang, F., Wagner, M. and Siddiqui, M.A.Q.  
 TITLE Structure, expression, and functional characterization of the mouse  
 Clp-1 gene  
 JOURNAL Gene 232 (1-2), 245-259 (2002)  
 MEDLINE 22114986  
 PUBMED 12119119  
 REFERENCE 2 (bases 1 to 3402)  
 AUTHORS Huang, F., Wagner, M. and Siddiqui, M.  
 TITLE Direct Submission  
 JOURNAL Submitted (19-MAR-2002) Anatomy and Cell Biology, State University  
 of New York Health Science Center, 450 Clarkson Avenue, Brooklyn,  
 NY 11203, USA  
 FEATURES  
 source  
 1..3402  
 /organism="Mus musculus"  
 /mol\_type="mRNA"  
 /strain="129/SvJ"  
 /db\_xref="taxon:10090"  
 1..3402  
 /gene="Clp1"  
 596..1666  
 /gene="Clp1"  
 /note="nuclear protein; Clp-1"  
 /codon\_start=1  
 /product="cardiac lineage protein 1"  
 /protein\_id="AA09026.1"  
 /db\_xref="GI:20135641"  
 /translation="MAEPLTEHOPOTNSCTGAAYVHEHTSERPSEAEVPEKED  
 SRWGRASLQSGRSGCEGEGELKHOPLQTNAPELSLSEKRGKQNGEDLSTGA  
 SPSAGEPMSSELVPGHDSATKQEPAPAGEEPMQOQRLGKKRRRRRSKKRH  
 MKPYKLTWEEKKKRDEKQSLASVRAEMFAGQVAPVNTQFLMDHDEEPLK  
 TGLYPRKAARKDDTSDDEPVEABEDGSDGMGDSGLFQROFSEYERYHAEVL  
 QNMSKOEILKEYLELEKLSRKEDENNRLRLESKRLGVDARVLELELDRLRAENT  
 QLTENELRQGERAPLSKFGD"  
 BASE COUNT 920 a 771 c 890 g 821 t  
 ORIGIN  
 Alignment Scores:  
 Pred. No.: 2,32e-75 Length: 3402  
 Score: 1587.50 Matches: 308  
 Percent Similarity: 88.86% Conservative: 11  
 Best Local Similarity: 85.79% Mismatches: 37  
 Query Match: 83.12% Indels: 3  
 DB: 10 Gaps: 2  
 US-09-972-758A-2 (1-359) x AY090614 (1-3402)  
 QY 1 MetAlaGluProPheLeuSerGluTyrGlnHisGlnProGlnInThSerAsnCyThrGly 20  
 DB 596 ATGGCCGAGCCCTCTTGCAGACATCAACACCAAGCTTCAAACTGTACAGGT 655  
 QY 21 AlaAlaAlaValGlnGlnGlnLeuAsnProGluArgProProGlyAlaGluGluArgVal 40  
 DB 656 GCTGCTGTTTCATTAAGACATACCTCTGAGCGCCCGCCAGCGCGAGGAGAGGGGTG 715  
 QY 41 ProGlulGluAspSerArgTrpGlnSerArgAlaPheProGlnLeuGlyValArgProGly 60  
 DB 716 CCGAAGGAGACAGTGTGGCAATGAGAGCGTCTTGCAATCGGTGCGGTCCAGGG 775  
 QY 61 ProGlulGlyGlySerLeuGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGln 80  
 DB 776 CAGGAGGAGAGAGGGGCGCTGAAGCACAGCTGCTCCCATTCGACAGCAATGCTGTCCA 835  
 QY 81 GluSerSerCysLeuArgGlnGlyGlyValGlnAsnGlyAspAspSerSerAlaGly 100  
 DB 836 GAATGAGCTCCCTGGAGAGGGCGAGAGGGCCAAATGGGAGAGCTTATTCACCTGGC 895  
 QY 101 GlyAspPheProProProAlaGluValGluProThrProGlnAlaGluLeuLeuAlaGln 120



```

Db      896 GGT---GCCTCCCGCGCGGAGGAGAGCCGATGTCAAGATCC-----CTCGTCGAG 946
Qy      121 ProCyH1eHspSerGluA1aSerLyLeuGluYAlaProAla1aGluYgluGlu 140
Db      947 CCAAGTCATGACTCGAGAGCCACCAAGCAGAGAGCTCTCCCGCGAGCGAGGCCA 1006
Qy      141 TTPGVLGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGln 160
Db      1007 TGGGAGACAGACAGACAGACAGCTGGGCGAAGAAAAACATCGAGAGCGCCCTCAAGAAAG 1066
Qy      161 LVAAGH1stProLyProTyTrrLyLeuThrTTPGVLGlnGlnGlnGlnGlnGln 180
Db      1067 AACGGCATTGGAAGCCCTTACTCAAGCTGACTGGAGAGAGAAAAAGTTGACGAG 1126
Qy      181 LysGlnSerLeuArgL1aSerArgL1aArgAlaGluMecPheAlaYsgLynProVal 200
Db      1127 AACGAGAGCTGGAGAGCTTGGCGGTTCCAGCCGAGATGTTCCCGCAAGGCCAGCAT 1186
Qy      201 AlaProTyRanThrThrglnPheLeuMecAspAspHisAspGlnGlnGlnGlnGln 220
Db      1187 GCGCCCTATPACACACAGCAGCTTCTCATGTGATGACACATCAGAGAGAGCTGATCTC 1246
Qy      221 LysThrGlyLeuTyRserLyArgAlaAlaAlaLysSerAspAspHisAspAsp 240
Db      1247 AAAACCGGCTTTTACCCCAAGCGGAGCGCGCAAAATCCAGACACACAGCATGAGGAT 1306
Qy      241 PheMetGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGln 260
Db      1307 TTTGTGGAGAGAGAGCTGTGTGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1366
Qy      261 GluPheLeuGlnArgAspPheSerGluThrTyRglnGlnGlnGlnGlnGlnGln 280
Db      1367 GAGTTTCTGAGAGAGAGCTTCTCGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1426
Qy      281 AsnMetSerLyGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGln 300
Db      1427 AACATGAGCAGAGAGAGAGCTCATCAAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1486
Qy      301 MetGlnAspGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGln 320
Db      1487 AAGGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1546
Qy      321 ArgValArgGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGln 340
Db      1547 CGAGTCGGGAGAGCTCGAGCTAGAGCTGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1606
Qy      341 ThrGlnAsnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGln 359
Db      1607 ACCGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1663

RESULT 11
AL731805      111002 bp      DNA      linear      ROD 21-JUN-2002
LOCUS      AL731805
DEFINITION      Mouse DNA sequence from clone Rp23-358E19 on chromosome 11,
complete sequence.
ACCESSION      AL731805
VERSION      AL731805.8 GI:21540125
KEYWORDS      HTG.
SOURCE      Mus musculus (house mouse)
ORGANISM      Mus musculus
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
1 (bases 1 to 111002)
Dunn,M.
Direct Submission
Submitted (21-JUN-2002) Wellcome Trust Sanger Institute, Hinxton,
Cambridgeshire, CB10 1SA, UK. E-mail enquiries:
humquery@sanger.ac.uk Clone requests: clonequest@sanger.ac.uk
on Jun 23, 2002 this sequence version replaced gi:21531483.
----- Genome Center
Center: Wellcome Trust Sanger Institute
Center code: SC

```

```

Web site: http://www.sanger.ac.uk
Contact: humquery@sanger.ac.uk
-----
During sequence assembly data is compared from overlapping clones.
Where differences are found these are annotated as variations
together with a note of the overlapping clone name. Note that the
variation annotation may not be found in the sequence submission
corresponding to the overlapping clone, as we submit sequences with
only a small overlap as described above.
This sequence was finished as follows unless otherwise noted: all
regions were either double-stranded or sequenced with an alternate
chemistry or covered by high quality data (i.e., phred quality >=
30); an attempt was made to resolve all sequencing problems, such
as compressions and repeats; all regions were covered by at least
one plasmid subclone or more than one M13 subclone; and the
assembly was confirmed by restriction digest. The following
abbreviations are used to associate primary accession numbers given
in the feature table with their source databases: Em: EMBL; Sw:
SWISSPROT; Tr: TrEMBL; Wp: WormPep; Info: Information on the WormPep
database can be found at
http://www.sanger.ac.uk/projects/C_elegans/wormpep_Rp23-358E19_is
from the RPI-23 Mouse PAC Library
constructed by the group of Pieter de Jong.
For further details see http://www.chori.org/bacpac/home.htm
VECTOR: pBACes.6.
Location/Qualifiers
1..111002
/organism="Mus musculus"
/mol_type="genomic DNA"
/db_xref="taxon:10090"
/chromosome="11"
/clone="RP23-358E19"
/clone_1b="RPI-23"
BASE COUNT      28051 a 27579 c 27135 g 28237 t
ORIGIN
Alignment Scores:
Pred. No.:      7,68e-74      Length:      111002
Score:      1587.50      Matches:      308
Percent Similarity:      88.86%      Conservative:      11
Best Local Similarity:      85.79%      Mismatches:      37
Query Match:      83.12%      Indels:      3
DB:      10      Gaps:      2

US-09-972-758a-2 (1-359) x AL731805 (1-111002)
Qy      1 MetAlaGluProPheLeuSerGluTyRglnH1sglnProGlnThrSeranCysThrGly 20
Db      91440 ATGGCCGAGCCACTCTTGACAGAACTCAACACAGCTCAACACTGACACTGACAGT 91499
Qy      21 AlaAlaValAlaGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGln 40
Db      91500 GCTGCTGTTCTCCATGAAAGAGATACCTGAGCGCCGCCCAAGCGCGAGAGAGCGGTG 91559
Qy      41 ProGlnGlnAspSerArgTrpGlnSerArgAlaPheProGlnGlnGlnGlnGlnGln 60
Db      91560 CCCAAGAGAGACAGTATGAGTGAATCGAGAGCGCTCTTCCAGTCCGTCGTCAGAG 91619
Qy      61 ProGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGln 80
Db      91620 CAGGAGGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 91679
Qy      81 GluSerSerCysLeuArgGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGln 100
Db      91680 GAATTGAGCTCCCTTGAGAGAGGCGAGAGAGGCGAGAGAGAGAGAGAGAGAGAG 91739
Qy      101 GluAspPheProProProAlaGlnValGluProThrProGlnValGlnGlnGlnGln 120
Db      91740 GGT---GCCTCCCGCGCGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 91790
Qy      121 ProCyH1aSpSerGlnA1aSerLyLeuGlnYAlaProAla1aGlnYgluGlnGln 140
Db      91791 CCAAGTCATGACTCGAGAGCCACCAAGCAGAGAGAGAGAGAGAGAGAGAGAGAGAGCA 91850

```



Qy	141	Tp1p1yngingninaargnleugllylsylyshisargtargaProserlylsy	160
Qy	91851	TGGGACACGACACAGACACAGCTGGCGAAGAAAAAATCATCGAGACGCCCTCAAGAGC	91911
Qy	161	LysArgHisrTplysProTyrTyrLysLeuThrTp1g1u1ylylsylyspheaspGlu	180
Db	91911	AAGCGCATTTGGAAAGCCCTCACTCAACAACCTGATGGAGAGAAAGAAAGTTTCAGACG	91970
Qy	181	LysGlnserLeuAargAlaserArgLlAargAlgluMetPheAlalysGlyInp-roval	200
Db	91971	AAGAGAGACCTCGAGGCTTCGGGGTTTCAGACCGAAGTGTTCGCAAGGGCCAGCCAGTT	92030
Qy	201	AlaProTyrAsnThrTrpGlnPheLeuMetAspSerHisAspGlnGluInguProAspLeu	220
Db	92031	GGCGCCTTATACACACACGACGACTTCTCATGATATACACCATCAGAGAGAGCCTGATCTC	92090
Qy	221	LysThrGlyLeuTyrSerLysAArgAlAlAlAlLysSerAspSerHisrAspAsp	240
Db	92091	AAAAACCGGCTTTATCCCAAGCGGGCAGCGCCCAATCCGACGACACACGATGTAGGAT	92150
Qy	241	PheMetGluGluGlyGlyGluGluAspGlyGlySerAspGlyMetGlyIysAspGlySer	260
Db	92151	TTTGTGGAGAGCTGGTGGAGAGAGAGAGAGAGCGCATG3CATGGAGGGAGACGGAGC	92210
Qy	261	GluPheLeuGlnAArgAspPheSerGluThrTyrGluArgTyrHisThrGlnSerLeuGln	280
Db	92211	GAGTTTCCGACGGGGACCTTCGAGACGATACGAGCGGTATCCACGCCGAGACCTTGACG	92270
Qy	281	AsnMetSerLysGlnGluLeuLlLysGluTyrLeuGluLeuGluTyrCysLeuSerArg	300
Db	92271	AACATGAGCAACGACGAGCTCATCAAAAGATACCTTGAGCTGGAGAAAGTCCCTCCCGC	92330
Qy	301	MetGluAspGluAsnAsnArgLeuArgLeuGlnSerLysArgLeuGlyIysAspAspAla	320
Db	92331	AAGAGAGACAAAATTAACGGCTCCGGCTGGAGAAACAAAGCGCTGGAGCGCTTCAGCGC	92390
Qy	321	ArgValAArgGluLeuGluLeuGluLeuAspArgLeuArgAlaGluAsnLeuGlnLeu	340
Db	92391	CGAGTGGGGGAGCTCGAGCTAGACTGAGACCGGCTGGCGCTGAGAACCTTCAGCTGCTG	92450
Qy	341	ThrGluAsnGluLeuHisArgGlnGlnGluArgAlaProLeuSerLysPheGlyAsp	359
Db	92451	ACCAGAGAACGAATCGACCGGCACAGAGAGCGGCTTTTCCAAAGTTCCGGCGAC	92507
RESULT 12			
AC120950/c			
LOCUS	AC120950	213625 bp	DNA linear HTG 19-NOV-2002
DEFINITION	Rattus norvegicus clone CH230-456h21, WORKING DRAFT SEQUENCE, 5 unordered pieces.		
ACCESSION	AC120950		
VERSION	AC120950.4	GI:25091749	
KEYWORDS	HTG, HTGS, PHASE1, HTGS_DRAFT, HTGS_FULLTOP.		
SOURCE	Rattus norvegicus (Norway rat)		
ORGANISM	Rattus norvegicus		
REFERENCE	1 (bases 1 to 213625)		
AUTHORS	Muzny,D,Marle, C, Metzker,M, Lee, A, Abramzon,S, Adams,C, Alder,J, Allen,C, Allen,H, Alsbrooks,S, Amin,A, Angulano,D, Anyalebechi,V, Aoyagi,A, Ayodeji,M, Baca,E, Baden,H, Baldwin,D, Bandaranaike,D, Barber,M, Barnstead,M, Benahmed,F, Biewald,K, Blair,J, Blankenburg,K, Blyth,P, Brown,M, Bryant,N, Buhay,C, Burch,P, Burrell,K, Calderon,E, Cardenas,V, Carter,K, Cavazos,I, Cesar,H, Center,A, Chacko,J, Chavez,D, Chen,G, Chen,R, Chen,Y, Chen,Z, Chu,J, Cleveland,C, Cockrell,R, Cox,C, Coyle,M, Crec,A, D'Souza,L, Davila,M,L, Davis,C, Davy-Carroll,L, De Anda,C, Deedrich,D, Delgado,O, Denson,S, Deramo,C, Ding,Y, Dinh,H, Diya,K, Draper,H, Dugan-Rocha,S, Dunn,A, Dublin,K, Duval,B, Eaves,K, Egan,A, Escotto,M, Eugene,C, Evans,C,A, Falle,T, Fan,G, Fernandez,S, Finley,M, Flagg,N, Forbes,L, Foster,M, Foster,P,		

TITLE  
 JOURNAL  
 REFERENCE  
 AUTHORS  
 TITLE  
 JOURNAL  
 REFERENCE  
 AUTHORS  
 TITLE  
 JOURNAL  
 COMMENT

Unpublished  
 2 (bases 1 to 213625)  
 Worley,K.C.  
 Direct Submission  
 Submitted (14-MAY-2002) Human Genome Sequencing Center, Department  
 of Molecular and Human Genetics, Baylor College of Medicine, One  
 Baylor Plaza, Houston, TX 77030, USA  
 3 (bases 1 to 213625)  
 Rat Genome Sequencing Consortium.  
 Direct Submission  
 Submitted (19-NOV-2002) Human Genome Sequencing Center, Department  
 of Molecular and Human Genetics, Baylor College of Medicine, One  
 Baylor Plaza, Houston, TX 77030, USA  
 On Nov 19, 2002 this sequence version replaced gi:23907976.  
 The sequence in this assembly is a combination of BAC based reads  
 and whole genome shotgun sequencing reads assembled using Atlas  
 (<http://www.hgsc.bcm.tmc.edu/projects/rat/>). Each contig described  
 in the feature table below represents a scaffold in the Atlas  
 assembly (a 'contig-scaffold'). Within each contig-scaffold,  
 individual sequence contigs are ordered and oriented, and separated  
 by sized gaps filled with Ns to the estimated size. The sequence  
 may extend beyond the ends of the clone and there may be sequence  
 contigs within a contig-scaffold that consist entirely of whole  
 genome shotgun sequence reads. Both end sequences and whole genome  
 shotgun sequence only contigs will be indicated in the feature  
 table.

----- Genome Center  
 Center: Baylor College of Medicine  
 Center code: BCM  
 Web site: <http://www.hgsc.bcm.tmc.edu/>  
 Contact: [hgsc-help@bcm.tmc.edu](mailto:hgsc-help@bcm.tmc.edu)  
 ----- Project Information  
 Center project name: GWRO  
 Center clone name: CH230-45CH21  
 ----- Summary Statistics  
 Assembly program: Phrap: statistics 0.990329  
 Consensus quality: 196084 bases at least Q40

Consensus quality: 197843 bases at least Q30  
 Consensus quality: 198926 bases at least Q20  
 Estimated insert size: 201713; sum-of-contrigs estimation  
 Quality coverage: 7x in Q20 bases; sum-of-contrigs estimation

NOTE: Estimated insert size may differ from sequence length  
 (see [http://www.hgsc.bcm.tmc.edu/docs/Genbank\\_draft\\_data.html](http://www.hgsc.bcm.tmc.edu/docs/Genbank_draft_data.html)).  
 NOTE: This is a 'working draft' sequence. It currently  
 consists of 5 contigs. The true order of the pieces  
 is not known and their order in this sequence record is  
 arbitrary. Gaps between the contigs are represented as  
 runs of N, but the exact sizes of the gaps are unknown.  
 This record will be updated with the finished sequence  
 as soon as it is available and the accession number will  
 be preserved.

1 58094: contig of 58094 bp in length  
 \* 58095 58194: gap of unknown length  
 \* 58195 139487: contig of 81303 bp in length  
 \* 139498 139597: gap of unknown length  
 \* 139598 208746: contig of 69149 bp in length  
 \* 208747 208846: gap of unknown length  
 \* 208847 210051: contig of 1205 bp in length  
 \* 210052 210151: gap of unknown length  
 \* 210152 213625: contig of 3474 bp in length.  
 Location/Qualifiers  
 1. 213625  
 /organism="Rattus norvegicus"  
 /mol\_type="genomic DNA"  
 /db\_xref="taxon:10116"  
 /clone="CH230-456H21"  
 1. 1190  
 /note="wgs end extension  
 clone\_end:Sp6"  
 47586 47846  
 /note="clone boundary  
 clone\_end:Sp6  
 site:  
 end sequence: B2240544"  
 45325 c 49186 g 48940 t 14232 others

BASE COUNT 51942 a 45325 c 49186 g 48940 t 14232 others  
 ORIGIN

Alignment Scores:  
 Pred. No.: 2,14e-73 Length: 213625  
 Score: 1584.50 Matches: 307  
 Percent Similarity: 88.58% Conservative: 11  
 Best Local Similarity: 85.52% Mismatches: 38  
 Query Match: 82.96% Indels: 3  
 DB: 2 Gaps: 2

US-09-972-758a-2 (1-359) x AC120950 (1-213625)

QY 1 MetAlaGluProPheLeuSerGluTyrGlnHisGlnProGlnThrSerAsnGlyThrGly 20  
 DB 119733 ATGGCTGAGCAGCTCTTGTGACAACTCAACACAGCTTCAACATGCACTTACAGGT 119674

QY 21 AAlaAlaValGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGln 40  
 DB 119673 GCGCTGCTTGTTCATAGAGGAGCAACTCTGAGCGCCGCCCAAGCGGAGGAGGAGG 119614

QY 41 ProGlnGluAspSerArgTyrGlnSerArgAlaPheProGlnLeuGlyGlyArgProGly 60  
 DB 119613 CCTAAGAGAGACAGTGTGAGCAATGAGAGCTCTTGCAGTCGGGTAGCCCTCCAGGG 119554

QY 61 ProGlnGlyGlnGlySerLeuGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGln 80  
 DB 119553 CAGGAGGAGACCGGGGCGCTGAAACACCGCTCCCGCATGCAAGATGCTCTCCA 119494

QY 81 GluSerSerCysLeuArgGlnGlyGlyGlyGlyGlnAsnGlyAspAspSerSerAlaGly 100  
 DB 119493 GAATTGAGCTGCTTGGAAAAAGCGGAGAGGCGCAAGATGGGAGGAGCTTATCCACTGGC 119434

QY 101 GlyAspPheProProAlaGluValGluProThrProGluAlaGluLeuLeuAlaGln 120

DB 119433 GGT---GCCTCGCTTCGCGGAGGAGGAGCCGATGTCAAGTCC-----CTCGTGCA 119383

QY 121 ProCysHisAspSerGluAlaSerIleuGlyAlaProAlaAlaGlyGlyGlnGlnGlu 140  
 DB 119382 CCAGGTCAATGACTCGGAGGCTCTTAAGCTGAGGCTCCGTGTGCTGAGATAGAGGAGCA 119323

QY 141 TTPGlyGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGln 160  
 DB 119322 TGGGAGCAGCAACAGAGACAGTGGGCAAGAAAAACCCGAGACGCCCTCCCAAG 119263

QY 161 LysArgHisTyrPheProTyrTyrIleuThrTyrPGLGlnGlyLysPheAspGlu 180  
 DB 119262 AGCGACATTTGAGAGCCCTATTACAGCTGACTTGGAGAGAGAGAGAGAGAGAGAGAG 119203

QY 181 LysGlnSerLeuArgAlaSerArgIleArgAlaGluMetPheAlaLysGlyGlnProVal 200  
 DB 119202 AAGCAAAAGCTTCGAGCTTCGAGAGTTCGAGCGAGATGTTGCCAAGGCGCCGGTT 119143

QY 201 AlaProTyrAsnThrThrGlnPheLeuMetAspAspHisAspGlnGluGluProAspLeu 220  
 DB 119142 GCGCCCTATTAACACACAGCAGGATTCCTCATGATGACCATGATCAGAGAGGCTGATCTC 119083

QY 221 LysThrGlyLeuTyrSerIleArgAlaAlaAlaLysSerAspAspThrSerAspAsp 240  
 DB 119082 AAAACCGGCTTTACCTTAAGCGGGAGCGCCCAATCCGAGACACACGAGTGGAGAT 119023

QY 241 PheMetGlnGlnGlyGlyGlyGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGln 260  
 DB 119022 TTCGTGAGAAAGACGTGTGAGAAAGCGAGGACAGAGATGGATGGAGGAGCGCACG 118963

QY 261 GluPheLeuGlnArgAspPheSerGluThrTyrGluArgTyrHisThrGluSerLeuGln 280  
 DB 118962 GAGTTCTTCACCGGAGCTTTCCAGACTTACGAGCGGTACACGCCCAAGAGCCCTGACG 118903

QY 281 AsnMetSerIleGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGln 300  
 DB 118902 AACATGAGCAGACAGAGAGCTCAACAAAGATCTGAGAGCTGAGAGAGGCTCTCCCGC 118843

QY 301 MetGluAspGluAsnAsnArgLeuArgLeuGlnSerIleArgLeuGlyGlyAspAspAla 320  
 DB 118842 AAGGAGAGCAAAATTAACCGGCTCGCTTGGAAGCAACGCGTTGGCTGCATAGCG 118783

QY 321 ArgValArgGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGln 340  
 DB 118782 CCAGTCCGGGAGCTGAGAGCTTACAGCTGCGACCGACGCGCGGAGAGAACCGCAAGCTGCTG 118723

QY 341 ThrGluAsnGlnLeuHisArgGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGln 360  
 DB 118722 ACCGAAAGCAACTGCACCGGAGAGGAGCGAGCGGCTCTTCCAAAGTTGGCGCAGC 118666

RESULT 13  
 BC022111 1488 bp mRNA linear ROD 16-APR-2003  
 LOCUS BC022111  
 DEFINITION Mus musculus cardiac lineage protein 1, mRNA (cdna clone MGC:36249  
 IMAGE:5067620), complete cds.  
 ACCESSION BC022111 GI:18381039  
 VERSION BC022111.1  
 KEYWORDS MGC.  
 SOURCE Mus musculus (house mouse)  
 ORGANISM Mus musculus  
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.  
 1 (bases 1 to 1488)  
 REFERENCE  
 AUTHORS Strausberg, R.L., Feingold, E.A., Grouse, L.H., Derge, J.G.,  
 Klausner, R.D., Collins, F.S., Wagner, L., Shenmen, C.M., Schuler, G.D.,  
 Altschul, S.F., Zeeberg, B., Buetow, K.H., Scheffer, C.F., Bhat, N.K.,  
 Hopkins, R.F., Jordan, H., Moore, T., Max, S.I., Wang, J., Hsieh, F.,  
 Diatchenko, L., Marusina, K., Farmer, A.A., Rubin, G.M., Hong, L.,  
 Stapleton, M., Soares, M.B., Bonaldo, M.F., Casavant, T.L.,  
 Scheetz, T.E., Brownstein, M.J., Usdin, T.B., Toshiyuki, S.,  
 Carrinci, P., Prange, C., Raha, S.S., Loughran, N.A., Peters, G.J.,  
 Abramson, R.D., Mullan, S.J., Bosak, S.A., McSwan, P.J.,  
 McKernan, K.J., Malek, J.A., Gunaratne, P.H., Richards, S.,

Worley, K.C., Hale, S., Garcia, A.M., Gay, L.J., Huliy, S.W., Villalon, D.K., Muzny, D.M., Sodergren, E.J., Lu, X., Gibbs, R.A., Valle, J., Helton, E., Kettner, M., Madan, A., Rodriguez, S., Sanchez, A., Whiting, M., Madan, A., Young, A.C., Shevchenko, Y., Bouffard, G.G., Blakesley, R.W., Touchman, J.W., Green, E.D., Dickson, M.C., Rodriguez, A.C., Grimwood, J., Schmutz, J., Myers, R.M., Butterfield, Y.S., Krzywinski, M.I., Skalska, U., Smalhus, D.E., Scherch, A., Schein, J.E., Jones, S.J., and Marra, M.A. Generation and initial analysis of more than 15,000 full-length human and mouse cDNA sequences  
Proc. Natl. Acad. Sci. U.S.A. 99 (26), 16899-16903 (2002)

JOURNAL MEDLINE 22388257  
PUBMED 12477932  
2 (bases 1 to 1488)  
Serausberg, R.  
Direct Submission  
Submitted (24-JAN-2002) National Institutes of Health, Mammalian Gene Collection (MGC), Cancer Genomics Office, National Cancer Institute, 31 Center Drive, Room 11A03, Bethesda, MD 20892-2590, USA  
NIH-MGC Project URL: <http://mgc.nci.nih.gov>  
Contact: MGC help desk  
Email: [cgabs-remail.nih.gov](mailto:cgabs-remail.nih.gov)  
Tissue Procurement: Gilbert Smith, Ph.D.  
cDNA Library Preparation: Life Technologies, Inc.  
cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LNLN)  
DNA Sequencing by: Genome Sequence Centre, BC Cancer Agency, Vancouver, BC, Canada  
[info@cgsc.bc.ca](mailto:info@cgsc.bc.ca)  
Steven Jones, Jennifer Amano, Ian Bosdet, Yaron Butterfield, Susana Chan, Readman Chiu, Chris Fjell, Erin Garland, Ran Guin, Leticia Hsiao, Martin Krzywinski, Reta Kutsche, Oliver Lee, Soo Sen Lee, Victor Ling, Carrie Mathewson, Candice McLeavy, Steven Ness, Pawan Pandoh, Anna-Lisa Prabhu, Parvaneh Saeedi, Jacqueline Schein, Duane Smalhus, Michael Smith, Lorraine Spence, Jeff Scott, Michael Thorne, Miranada Tsai, Natasja van den Bosch, Jill Vardy, George Yang, Scott Zuyderduyn, Marco Marra.

Clone distribution: MGC clone distribution information can be found through the I.M.A.G.E. Consortium/LNLN at: <http://image.lnl.gov>  
Series: IRAX Plate: 57 Row: a Column: 15  
This clone was selected for full length sequencing because it passed the following selection criteria: Hexamer frequency ORF analysis, similarity but not identity to protein.

FEATURES  
source  
1..1488  
/organism="Mus musculus"  
/mol\_type="mRNA"  
/strain="CZECH II"  
/db\_xref="taxon:10090"  
/clone="MGC:36249 IMAGE:5067620"  
/tissue\_type="Mammary tumor metastasized to lung. Tumor arose spontaneously from a senescent normal mammary (clonal) outgrowth infected with the virus MMTV."  
/clone\_1db="NCI CGAP Lu29"  
/lab\_host="DH10B"  
/note="Vector: pCMV-SPORT6"  
1..1488  
/gene="C1p1-pending"  
/note="synonyms: H1s1, CLP-1"  
/db\_xref="locusid:192231"  
/db\_xref="mgi:2385923"  
448..1188  
/codon\_start=1  
/product="C1p1-pending protein"  
/protein\_id="AAH2111.1"  
/db\_xref="GI:18381040"  
/db\_xref="locusid:192231"  
/translation="MSESLVDPGHSATKQFAPAAGSEPMGQROGLKKHRRP SKKRHKPYRKLTWEKKKPKDEKQSIKASRAVRAMKQAPAPATTTDPLDDHDQ EEDPLKGLPKRAAKSDSDTSDEDFVEEAGEEDGSDGSGSEFLDPSFETTER VHAESLONMSKOEILKEYLEIKLSKKEBNRLRLRESKRLGIVDAVLELLEIDR LRAENILQLTTENELHRCQERAPLSKFGD"

BASE COUNT 385 a 371 c 449 g 283 t  
ORIGIN  
Alignment Scores:  
Pred. No.: 3 42e-75 Length: 1488  
Score: 1577.50 Matches: 305  
Percent Similarity: 88.86 Conservative: 14  
Best Local Similarity: 84.96 Mismatches: 37  
Query Match: 82.59 Indels: 3  
DB: 10 Gaps: 2  
US-09-972-758a-2 (1-359) x BC022111 (1-1488)  
QY 1 MetAlaGluProPheLeuSerGluTyrGlnHisGlnProGlnThrSerAnCySerThrGly 20  
Db 118 GTCACACACCACTCTTGACAGAACATCAACACGACCTCAACTGACATGACAGGT 177  
QY 21 AAlaAlaValGlnGlnGluLeuAsnProGluValProProGlyValAglGluVal 40  
Db 178 GCTGTGTGTTCATGAGCATACCTCTGAGCGCCCAAGCGCGAGGAAACGGGTG 237  
QY 41 ProGluGluAspSerArgTrpGlnSerArgAlaPheProGlnLeuGlyValArgProGly 60  
Db 238 CCAAGAGAGACATGATGTCGACATGACAGAGCTCTTGCAATCCGATCCGATCCAGGG 297  
QY 61 ProGluGlyGluGlySerLeuGluSerGlnProProProLeuGlnThrGlnAlaCyPro 80  
Db 298 CAGAGAGAGAGAGGGGGCTGAAAGCACAGCTGCCCAATTCGACAGCATGCTGTCCA 357  
QY 81 GluSerSerCysLeuArgGluGlyGlyValGlnAsnGlyAspAspSerAlaGly 100  
Db 358 GAATGAGCTCCCTGAGAGAGCGGAGAGCGGCAAAAGGAGAGACTTATTCACCTGGC 417  
QY 101 GlyAspPheProProProAlaGluValGluProThrProGluAlaGluLeuLeuAlaGln 120  
Db 418 GGT---GCTCCCGCGCGGAGGAGAGCCGATCTCAAGTCC-----CTCGTCCAG 468  
QY 121 ProCysHisAspSerGluAlaSerGlyLeuGlyAlaProAlaAlaGlyGlyGluGlu 140  
Db 469 CCAAGCATGACTCGAGGCGCACCAAGAGAGGCTCCGCGCGGAGCGGAGGAGGCCA 528  
QY 141 TrpGlyGlnGlnGlnArgGlnLeuGlyValGlyValGlnArgArgProSerGlyVal 160  
Db 529 TGGGACACACACAGAGACAGCTGGGCAAGAAAACATCGAGAGCGCCCTCAAGAAAG 588  
QY 161 LysArgHisTrpLysProTyrTyrLysLeuThrTrpGluGluLysLysPheAspGlu 180  
Db 589 AAGCGGCAATTGAGAGCCCTACTACAGCTGAGTGGAGAGAGAAAAGTTGACGAG 648  
QY 181 LysGlnSerLeuArgAlaSerArgIleArgAlaGluMetPheAlaLysGlyGlnProVal 200  
Db 649 AAGCAGAGCTCGAGACTTCGCGGTTCAGAGCGAGATTTGGCCAAAGGCGACGAGTT 708  
QY 201 AlaProTyrAsnThrThrGlnPheLeuMetAspAspHisAspGlnGluGluProAspLeu 220  
Db 709 GCGCCTATTAACACACGAGTTCCTCATGATGACCAAGATCAGAGAGAGCGCTGATCTC 768  
QY 221 LysThrGlyLeuTyrSerLysArgAlaAlaAlaLysSerAspAspThrSerAspAsp 240  
Db 769 AAAACCGGCTTTTACCCCAAGCGGAGCGGCAATTCGAGACACCAAGCATGAGAT 828  
QY 241 PheMetGluGluGlyGlyGluGluAspGlyLysSerAspGlyMetGlyGlyLysArg 260  
Db 829 TTTGTGAAGAAAGCTGTGAGAAAGCGAGCGAGATGCGATGGAGGAGGAGCGGAC 888  
QY 261 GluPheLeuGlnArgAspPheSerGluThrTyrGluArgTyrHisThrGluSerLeuGln 280  
Db 889 GAGTTCTCAGCGGAGACTTCTCGAGAGCGTACGAGCGGTACCAACCGAGAGCGCTCGAG 948  
QY 281 AsnMetSerLysGlnGluLeuLysGluTyrLeuGluLeuGluLysCysLeuSerArg 300  
Db 949 AACATGAGCAAGAGAGAGCTCATCAAGAGTACCTGAGAGGAGAGAGTCTCTCCGC 1008

REFERENCE	AUTHORS	LOCUS	DEFINITION	ACCESSION	VERSION	KEYWORDS	SOURCE	ORGANISM
Oy	301 MetGluApBgJuaAsnAsnAryLeuArgLeuGluGluSerLysAArgLeuGluGJGJVApsAspA 320							
Db	1009 AAGAGGACGCAAAATTAACCGCTGGCTGGAAACCAAGCGGCTGGAGGCTGCACGCC 1066							
Oy	321 ArgValArgJuleGluGluGluGluLeuAspAryLeuArgAaGluAaGluAsnLeuGluLeu 340							
Db	1069 CGAGTGGCGGAGCTTCGACCTAGAGCTGGACCGGCTGCAGCACTCCAGACTGCTG 1128							
Oy	341 ThrGluGluGluLeuHisAryGluGluGluGluArgAaLProLeuSerLysPheGluAAsp 359							
Db	1129 ACCGAGAACGAACTGCACCGGACGACGAGGACGAGCGCTCTTCCAACTTGGCGCAC 1185							
AC136172/c								
LOCUS	AC136172	11803 bp	DNA	linear	HTG 30-OCT-2002			
DEFINITION	Rattus norvegicus clone CH230-226J2, *** SEQUENCING IN PROGRESS							
ACCESSION	AC136172							
VERSION	AC136172.1 GI:24421642							
KEYWORDS	HTG, HTGS, PHASE1							
SOURCE	Rattus norvegicus (Norway rat)							
ORGANISM	Rattus norvegicus Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Rattus.							
REFERENCE	1 (bases 1 to 11803)							
AUTHORS	Muzny, D., Maric, Metzker, M., Lee, Abramson, S., Adams, C., Alder, J., Allen, C., Allen, H., Albrooks, S., Amin, A., Angiano, D., Anylebech, V., Aoyagi, A., Ayodeji, M., Baca, E., Baden, H., Baldwin, D., Bandaranaike, D., Barber, M., Barnstead, M., Benahmed, F., Bitwalo, K., Blair, J., Blankenburg, K., Blyth, P., Brown, M., Bryant, N., Bunay, C., Burch, P., Burrell, K., Calderon, E., Cardenas, J., Carter, K., Cavazos, I., Cesaari, H., Center, A., Chacko, J., Chavez, D., Chen, G., Chen, R., Chen, Y., Chen, Z., Chu, J., Cleveland, C., Cockrell, R., Cox, C., Coyle, M., Cree, A., D'Souza, M., Davis, M. L., Davis, C., Davy-Carrillo, L., De And, C., Dedrich, D., Delgado, O., Denon, S., Detamo, C., Ding, Y., Dinh, H., Divya, K., Draper, H., Dugan-Rocha, S., Dunn, A., Durbin, K., Duval, B., Eaves, K., Egan, A., Escotto, M., Eugene, C., Evans, C. A., Falls, T., Fan, G., Fernandez, S., Finley, M., Flagg, N., Forbes, L., Foster, M., Foster, P., Fraser, C. M., Gabisi, A., Ganta, R., Garcia, A., Garner, T., Garra, M., Gegeorgis, E., Geer, K., Gill, R., Grady, M., Guerra, W., Guetara, W., Gunaratne, P., Haaland, W., Hamill, C., Hamilton, C., Hamilton, K., Harvey, Y., Havlak, P., Hawes, A., Henderson, N., Hernandez, J., Hernandez, S., Hines, S., Hladun, S. L., Hodgson, A., Hognes, M., Hollins, B., Howells, S., Hulik, S., Hume, J., Idlebird, D., Jackson, A., Jackson, L., Jacob, L., Jiang, H., Johnson, B., Johnson, R., Jolivet, A., Karpathy, S., Kelly, S., Khan, Z., King, L., Kovar, C., Kowals, C., Krafic, C. L., Lebow, H., Levan, J., Lewis, L., Li, Z., Liu, J., Liu, J., Liu, W., Liu, Y., London, P., Longacre, S., Lopez, J., Lorenzuewa, L., Louised, H., Lozada, R. J., Lu, X., Ma, J., Maheshwari, M., Mahindartne, M., Mahmood, M., Malloy, K., Mangun, A., Mangum, B., Mapua, P., Martin, K., Martin, R., Martinez, E., Mahoney, S., McLeod, M., McNeill, T., Meenen, E., Milosevic, A., Miner, G., Mijic, E., Montemayor, J., Moore, S., Morgan, M., Morris, K., Morris, S., Munkidasa, M., Murphy, M., Nair, L., Nankervis, C., Neal, D., Newton, N., Nguyen, N., Norris, S., Nwokelemeh, O., Okunou, G., Olarinmusaogun, A., Pal, S., Parks, K., Pasternak, S., Paul, H., Perez, A., Perez, L., Pfannkuch, C., Plopper, F., Poidexter, A., Popovic, D., Primus, E., Pu, L., Puazo, M., Quitor, J., Rachlin, E., Reeves, K., Regier, M. A., Reigh, R., Reilly, B., Reilly, M., Ren, Y., Reuter, M., Richards, S., Riggs, F., Rives, C., Rodkey, T., Rojas, A., Rose, M., Rose, R., Ruiz, S. J., Sanders, W., Savery, G., Scherer, S., Scott, G., Shlتمان, S., Shen, H., Shetty, J., Shvartbeyn, A., Slason, I., Sitter, C. D., Smaj, D., Sneed, A., Soderren, E., Song, X.-Z., Sorelle, R., Soes, J., Steidle, M., Strong, R., Sutton, A., Svatek, A., Tabori, P., Taylor, C., Taylor, T., Thomas, N., Thomas, D., Tingey, A., Trujes, Z., Usmani, K., Valas, R., Vers, V., Villasaana, S., Waldron, L., Walker, B., Wang, J., Wang, Q., Wang, S., Warren, J., Warren, R., Wei, X., White, P., Williams, G., Willson, R., Wleczk, R., Wooden, H., Worley, K., Wright, D., Wright, R., Wu, J., Yakub, S., Yen, J., Yoon, V., Yoon, Y., Yu, F., Zhang, Y., Zhou, J., Zhou, X., Zhao, S., Dunn, D., von Niederhausern, A., Weiss, R., Smith, D. R.,							

```

JOURNAL TITLE
REFERENCE AUTHORS
JOURNAL TITLE
COMMENT
Holt, R. A., Smith, H. O., Weinstein, G. and Gibbs, R. A.
Direct Submission
Unpublished
2 (bases 1 to 11803)
Rat Genome Sequencing Consortium.
Direct Submission
Submitted (30-OCT-2002) Human Genome Sequencing Center, Department
of Molecular and Human Genetics, Baylor College of Medicine, One
Baylor Plaza, Houston, TX 77030, USA
-----
Center: Baylor College of Medicine
Center code: BCM
Web site: http://www.hgsc.bcm.tmc.edu/
Contact: hgsc-help@bcm.tmc.edu
-----
Project Information
Center project name: KCV1
Center Clone name: CH230-22692
-----
Sequencing vector: Plasmid;
Chemistry: Dye-terminator Big Dye; 100% of reads
Assembly program: Phrap; version 0.990329
Consensus quality: 67362 bases at least Q40
Consensus quality: 72560 bases at least Q30
Consensus quality: 77367 bases at least Q20
-----
NOTE: Estimated insert size may differ from sequence length
(see http://www.hgsc.bcm.tmc.edu/docs/genbank_direct_data.html)
NOTE: This is a 'working draft' sequence. It currently
consists of 53 contigs. The true order of the pieces
is not known and their order in this sequence record is
arbitrary. Gaps between the contigs are represented as
runs of N, but the exact sizes of the gaps are unknown.
This record will be updated with the finished sequence
as soon as it is available and the accession number will
be preserved.
-----
1
1245 1244: contig of 1244 bp in length
1345 1344: gap of unknown length
2653 1309: contig of 1309 bp in length
2654 2753: gap of unknown length
2754 4274: contig of 1521 bp in length
4275 4374: gap of unknown length
4375 5978: contig of 1604 bp in length
5979 6078: gap of unknown length
6079 7135: contig of 1057 bp in length
7136 7235: gap of unknown length
7236 8520: contig of 1285 bp in length
8521 8620: gap of unknown length
8621 10172: contig of 1552 bp in length
10173 10272: gap of unknown length
10273 11379: contig of 1107 bp in length
11380 11479: gap of unknown length
11480 12576: contig of 1097 bp in length
12577 12676: gap of unknown length
12677 13863: contig of 1187 bp in length
13864 13963: gap of unknown length
13964 14987: contig of 1024 bp in length
14988 15087: gap of unknown length
15088 16573: contig of 1486 bp in length
16574 16673: gap of unknown length
16674 17896: contig of 1223 bp in length
17897 17996: gap of unknown length
17997 19475: contig of 1479 bp in length
19476 19575: gap of unknown length
19576 20884: contig of 1309 bp in length
20885 20984: gap of unknown length
20985 22487: contig of 1503 bp in length
22488 22587: gap of unknown length
22588 24747: contig of 2160 bp in length
24748 24847: gap of unknown length
24848 26172: contig of 1325 bp in length
26173 26272: gap of unknown length
26273 27802: contig of 1530 bp in length
27803 27902: gap of unknown length

```

```

* 27903 29918: contig of 2016 bp in length
* 29919 30018: gap of unknown length
* 30019 31165: contig of 1147 bp in length
* 31166 31265: gap of unknown length
* 31266 32793: contig of 1528 bp in length
* 32794 32893: gap of unknown length
* 32894 34401: contig of 1508 bp in length
* 34402 34501: gap of unknown length
* 34502 35644: contig of 1143 bp in length
* 35645 35744: gap of unknown length
* 35745 36968: contig of 1224 bp in length
* 36969 37068: gap of unknown length
* 37069 38349: contig of 1281 bp in length
* 38350 38449: gap of unknown length
* 38450 41409: contig of 2960 bp in length
* 41410 41509: gap of unknown length
* 41510 43066: contig of 1557 bp in length
* 43067 43166: gap of unknown length
* 43167 44827: contig of 1661 bp in length
* 44828 44927: gap of unknown length
* 44929 46592: contig of 1665 bp in length
* 46593 46692: gap of unknown length
* 46693 49255: contig of 2563 bp in length
* 49256 49355: gap of unknown length
* 49356 51247: contig of 1892 bp in length
* 51248 51347: gap of unknown length
* 51348 53114: contig of 1767 bp in length
* 53115 53214: gap of unknown length
* 53215 55332: contig of 2118 bp in length
* 55333 55432: gap of unknown length
* 55433 57427: contig of 1995 bp in length
* 57428 57527: gap of unknown length
* 57528 59606: contig of 2079 bp in length
* 59607 59706: gap of unknown length
* 59707 61138: contig of 1432 bp in length
* 61139 61238: gap of unknown length
* 61239 63710: contig of 2472 bp in length
* 63711 63810: gap of unknown length
* 63811 66371: contig of 2561 bp in length
* 66372 66471: gap of unknown length
* 66472 70029: contig of 3558 bp in length
* 70030 70129: gap of unknown length
* 70130 72934: contig of 2805 bp in length
* 72935 73034: gap of unknown length
* 73035 76024: contig of 2990 bp in length
* 76025 76124: gap of unknown length
* 76125 79346: contig of 3222 bp in length
* 79347 79446: gap of unknown length
* 79447 82950: contig of 3504 bp in length
* 82951 83050: gap of unknown length
* 83051 86116: contig of 3066 bp in length
* 86117 86216: gap of unknown length
* 86217 89119: contig of 2903 bp in length
* 89120 89219: gap of unknown length
* 89220 92494: contig of 3275 bp in length
* 92495 92594: gap of unknown length
* 92595 94419: contig of 1825 bp in length
* 94420 94519: gap of unknown length
* 94520 98048: contig of 3529 bp in length
* 98049 98148: gap of unknown length
* 98149 101599: contig of 3451 bp in length
* 101600 101699: gap of unknown length
* 101700 105918: contig of 4219 bp in length
* 105919 106018: gap of unknown length
* 106019 108449: contig of 2431 bp in length
* 108450 108549: gap of unknown length
* 108550 111803: contig of 3254 bp in length.

```

## FEATURES

Alignment Scores: 3.58e-36 Length: 111803  
 Pred. No.: 875.00 Matches: 177  
 Score: 94.74% Conservative: 3  
 Percent Similarity: Location/Qualifiers

Best Local Similarity: 93.16% Mismatches: 10  
 Query Match: 45.81% Indels: 2  
 DB: 2 Gaps: 0

US-09-972-758a-2 (1-359) x AC136172 (1-111803)

```

QY 170 LeuThrTTPGLuGLuLyLySphEaPGLuLyGlnSerLeuArgAlaSerArgIle 189
DB 14987 CTGACTTGAG-GAGAAAGAAAAGTTCCAGCAGAAAGCAAGCTGCGACTTCGACGTT 14929

QY 190 ArgAlaGluMetPheAlaIysGlnProValAlaProTyrAsnThrGlnPheLeu 209
DB 14928 CGAGCCGAGATGTC-GCCAAAGGCGACCGGTTGGCCCTATTAACACCGACGATTCGAG 14870

QY 210 MetaSpAspHisAspGlnGluProAspLeuLysThrGlyLeuTyrSerIysArgAla 229
DB 14869 ATGGATGACCATGATGAGAGAGAGAGCTGATCTCAAAACCGGCTTTACCTTAAGGGGCA 14810

QY 230 AlaAlaIysSerAspAspThrSerAspAspAspPheMetGluGluGlyGluIleAsp 249
DB 14809 GCGGCCAAATCCGACGACACACAGCATGAGATTTCGTGAAGAGCTGTGAGGAAGAC 14750

QY 250 GlyGlySerAspGlyMetGlyIysAspGlySerGluPheLeuGlnArgAspPheSerGlu 269
DB 14749 GAGGCGAGGATGCGATGGAGGAGGAGCGAGCATTTCTGACGCGACTTTTCGAG 14690

QY 270 ThrTyrGluArgTyrHisThrGluSerLeuGlnAsnMetSerIysGlnIleuIleLys 289
DB 14689 ACTTACGACGGTATCCACCCGACGACCTGCGAATATAGCAGAGAGCTCATCAA 14630

QY 290 GluTyrLeuGluLeuGluLysCysLeuSerArgMetGluAspGluAsnAsnArgLeuArg 309
DB 14629 GAGTACTCGAGCTGAGAGAAGTGCTCTCCCGCAAGAGAGACGAAATAACCGGCTGCGG 14570

QY 310 LeuGluSerIysArgLeuGlyGlyIysAspAlaArgValArgGluLeuGluIleuLys 329
DB 14569 TTGGAAGAGCAACGCGTGGGTGGCTGATGGCGAGTCCGGAGCTGAGAGCTGAGAGCTG 14510

QY 330 AspArgLeuArgAlaGluAsnLeuGlnLeuLeuThrGluAsnGluLeuHisArgGlnGln 349
DB 14509 GACCGACTCGCGCGCGAGAACCGCCAGCTGCTGACCGAAGAACGAACTGCACCGGCGACG 14450

QY 350 GluArgAlaProLeuSerIysPheGlyAsp 359
DB 14449 GAGCGAGCGCTCTCCAAAGTTCCGCGAC 14420

RESULT 15
LOCUS BD058159 523 bp DNA linear PAT 27-AUG-2002
DEFINITION Secreted expressed sequence tags (ESTs).
ACCESSION BD058159
VERSION BD058159.1 GI:22603765
KEYWORDS JP 2001519666-A/14.
SOURCE Zea mays
ORGANISM Zea mays
Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae; PACAD
clade; Panicoideae; Andropogoneae; Zea.
REFERENCE 1 (bases 1 to 523)
AUTHORS Jacobs,K., Mccoy,J.M., Lavallie,E.R., Racie,L.A., Merberg,D.,
Treacy,M., Spaulding,V. and Agostino,M.J.
Secreted expressed sequence tags (ESTs).
Patent: JP 2001519666-A 14 23-OCT-2001;
JOURNAL GENETICS INSTITUTE INC
COMMENT PN JP 2001519666-A/14
PD 23-OCT-2001
PF 10-APR-1998 JP 1998543068
PR 10-APR-1997 US 08/835913
PI KENNETH JACOBS, JOHN M MCCOY, EDWARD R LAVALLIE, LISA A RACIE, PI
DAVID MERBERG,
PI MAURICE TREACY,VIKKI SPAULDING,MICHAEL J AGOSTINO PC
C12N15/12,C12N5/10,C07K14/47,C12Q1/68,A61K38/17 CC Strandedness:
Double;

```

CC Topology: Linear;  
FH Key Location/Qualifiers.  
FEATURES  
Source 1..523  
/organism="Zea mays"  
/mol\_type="genomic DNA"  
/db\_xref="taxon:4577"

BASE COUNT 112 a 149 c 153 g 108 t 1 others  
ORIGIN

## Alignment Scores:

Pred. No.:	5.32e-30	Length:	523
Score:	714.00	Matches:	131
Percent Similarity:	99.26%	Conservative:	3
Best Local Similarity:	97.04%	Mismatches:	1
Query Match:	37.38%	Indels:	0
DB:	6	Gaps:	0

US-09-972-758a-2 (1-359) x BD058159 (1-523)

```
QY 1 MetAlaGluProPheLeuSerGluTyrGlnHisGlnProGlnThrSerAsnCystrGly 20
   |||||
Db 104 ATGCCGAGGCACTTCTGTGAGATATCAACACAGCCTCAACTAGCACTGTACAGT 163
   |||||
QY 21 AlaAlaIaValGlnGluLeuAsnProGluArgProGlyAlaGluGluArgVal 40
   |||||
Db 164 GCTGCTGCTGTCCAGAAAGAGCTGAACCTGAGCGCCCCAGCGCGAGAGCGGGTG 223
   |||||
QY 41 ProGluGluAspSerArgTArgInserArgAlaPheProGlnLeuGlyArgProGly 60
   |||||
Db 224 CCCGAGGAGGACAGTAGGTGGCAATCGAGACGTTCCCAAGTTGGGTGGCCGCGGG 283
   |||||
QY 61 ProGluGlyGluGlySerLeuGluSerGlnProProProLeuGlnThrGlnAlaCysPro 80
   |||||
Db 284 CCGGAGGAGGAGGAGGAGCCTGGAATCCCAACACACTCTCTTGCAAGACCCAGGCTGTCCA 343
   |||||
QY 81 GluSerSerCysLeuArgGluGlyGluGlyGlnAsnGlyAspAspSerSerAlaGly 100
   |||||
Db 344 GAATCTAGCTGCTGAGAGAGAGGCGGAGAAAGGCGCAGAAATGGGAGCGACTGCTCGCTGGC 403
   |||||
QY 101 GlyAspPheProProProAlaGluValGluProThrProGluAlaGluLeuLeuAlaGln 120
   |||||
Db 404 GGGGACTTCCGCGCGCGGAGAGAGGAAACGACGCGCGGCGGCGGCTGCTGCCAG 463
   |||||
QY 121 ProCysHisAspSerGluAlaSerLysLeuGlyValaProAlaAla 135
   |||||
Db 464 CTTGTGATGACTCCGAGGCCAGTAAGATGGGTGCAAGGCTCTCT 508
   |||||
```

Search completed: February 5, 2004, 15:09:45  
Job time : 4896 secs

GenCore version 5.1.6  
Copyright (c) 1993 - 2004 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: February 5, 2004, 13:31:10 ; Search time 92 Seconds  
(without alignments)  
817,046 Million cell updates/sec

Title: US-09-972-758a-2

Perfect score: 1910  
Sequence: 1 MAEPFLSEYQHQPOTSNCTG.....LTENELHROERAPLSKFD 359

Scoring table: BLOSUM62  
Gapop 10.0 , Gapext 0.5

Searched: 801455 seqs, 209382283 residues

Total number of hits satisfying chosen parameters: 801455

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : Published Applications\_AA:\*

- 1: /cgn2\_6/ptodata/2/pubpaa/US07\_PUBCOMB.pep:\*
- 2: /cgn2\_6/ptodata/2/pubpaa/PCT\_NEW\_PUB.pep:\*
- 3: /cgn2\_6/ptodata/2/pubpaa/US06\_NEW\_PUB.pep:\*
- 4: /cgn2\_6/ptodata/2/pubpaa/US06\_PUBCOMB.pep:\*
- 5: /cgn2\_6/ptodata/2/pubpaa/US07\_NEW\_PUB.pep:\*
- 6: /cgn2\_6/ptodata/2/pubpaa/PCTUS\_PUBCOMB.pep:\*
- 7: /cgn2\_6/ptodata/2/pubpaa/US08\_NEW\_PUB.pep:\*
- 8: /cgn2\_6/ptodata/2/pubpaa/US08\_PUBCOMB.pep:\*
- 9: /cgn2\_6/ptodata/2/pubpaa/US09A\_PUBCOMB.pep:\*
- 10: /cgn2\_6/ptodata/2/pubpaa/US09B\_PUBCOMB.pep:\*
- 11: /cgn2\_6/ptodata/2/pubpaa/US09C\_PUBCOMB.pep:\*
- 12: /cgn2\_6/ptodata/2/pubpaa/US09\_NEW\_PUB.pep:\*
- 13: /cgn2\_6/ptodata/2/pubpaa/US10A\_PUBCOMB.pep:\*
- 14: /cgn2\_6/ptodata/2/pubpaa/US10B\_PUBCOMB.pep:\*
- 15: /cgn2\_6/ptodata/2/pubpaa/US10C\_PUBCOMB.pep:\*
- 16: /cgn2\_6/ptodata/2/pubpaa/US10\_NEW\_PUB.pep:\*
- 17: /cgn2\_6/ptodata/2/pubpaa/US60\_NEW\_PUB.pep:\*
- 18: /cgn2\_6/ptodata/2/pubpaa/US60\_PUBCOMB.pep:\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

#### SUMMARIES

Result No.	Score	Query Match	Length	DB	ID	Description
1	1910	100.0	359	9	US-09-745-763-34	Sequence 34, Appl
2	1910	100.0	359	10	US-09-972-758-2	Sequence 2, Appl
3	1910	100.0	359	15	US-10-205-823-172	Sequence 172, Appl
4	545.5	28.6	286	12	US-10-094-749-2862	Sequence 2862, Appl
5	134.5	7.0	546	10	US-09-976-740-44	Sequence 44, Appl
6	134.5	7.0	546	12	US-10-616-187-44	Sequence 44, Appl
7	134.5	7.0	546	14	US-10-023-529-44	Sequence 44, Appl
8	134.5	7.0	546	14	US-10-023-523-44	Sequence 44, Appl
9	130	6.8	611	9	US-09-216-393-81	Sequence 81, Appl
10	130	6.8	611	12	US-10-321-856-81	Sequence 81, Appl
11	128.5	6.7	804	12	US-10-108-2604-1161	Sequence 4161, Appl
12	128	6.7	2476	11	US-09-824-574-7	Sequence 7, Appl
13	126.5	6.6	722	12	US-10-161-051-193	Sequence 193, Appl
14	125	6.5	333	12	US-10-259-194A-320	Sequence 320, Appl
15	125	6.5	383	12	US-10-374-780A-2854	Sequence 2854, Appl

16	125	6.5	650	12	US-10-374-780A-514	Sequence 514, App
17	124	6.5	489	9	US-09-876-889-350	Sequence 350, App
18	123.5	6.5	557	9	US-09-962-055-5	Sequence 5, Appl
19	123.5	6.5	557	10	US-09-976-740-5	Sequence 5, Appl
20	123.5	6.5	557	12	US-10-616-187-5	Sequence 5, Appl
21	123.5	6.5	557	14	US-10-023-529-5	Sequence 5, Appl
22	123.5	6.5	557	14	US-10-023-523-5	Sequence 5, Appl
23	123.5	6.5	557	12	US-10-104-047-3378	Sequence 3378, Ap
24	123.5	6.5	5008	12	US-10-051-874-166	Sequence 166, App
25	121.5	6.4	342	12	US-10-317-833-118	Sequence 118, App
26	121.5	6.4	1974	12	US-10-369-493-6395	Sequence 6395, Ap
27	121	6.3	537	12	US-10-051-874-138	Sequence 138, App
28	121	6.3	538	9	US-09-827-822-9	Sequence 9, Appl
29	120	6.3	427	15	US-10-050-704-273	Sequence 273, App
30	120	6.3	483	15	US-10-050-704-272	Sequence 272, App
31	120	6.3	911	12	US-10-237-496-52	Sequence 52, Appl
32	120	6.3	911	12	US-10-242-505-52	Sequence 52, Appl
33	120	6.3	911	12	US-10-242-505-52	Sequence 52, Appl
34	120	6.3	911	12	US-10-242-574-52	Sequence 52, Appl
35	120	6.3	911	12	US-10-243-261-52	Sequence 52, Appl
36	120	6.3	911	12	US-10-243-282-52	Sequence 52, Appl
37	120	6.3	911	12	US-10-243-402-52	Sequence 52, Appl
38	120	6.3	911	12	US-10-243-431-52	Sequence 52, Appl
39	120	6.3	911	12	US-10-245-164-52	Sequence 52, Appl
40	120	6.3	911	12	US-10-245-164-52	Sequence 52, Appl
41	120	6.3	911	12	US-10-197-942-52	Sequence 52, Appl
42	120	6.3	911	12	US-10-238-196-52	Sequence 52, Appl
43	120	6.3	911	12	US-10-245-013-52	Sequence 52, Appl
44	120	6.3	911	12	US-10-295-027-260	Sequence 260, App
45	120	6.3	911	15	US-10-245-103-52	Sequence 52, Appl

#### ALIGNMENTS

RESULT 1  
US-09-745-763-34  
Sequence 34, Application J97/09745763  
Patent No. US20020065394A1  
GENERAL INFORMATION:

APPLICANT: Jacobs, Kenneth

Mccoy, John M.

Lavallee, Edward R.

Collins-Racie, Lisa A.

Evans, Cheryl

Merberg, David

Ireacy, Maurice

Spaulding, Vikki

TITLE OF INVENTION: SECRETED PROTEINS AND POLYNUCLEOTIDES

ENCODING THEM

NUMBER OF SEQUENCES: 219

CORRESPONDENCE ADDRESS:

ADDRESSEE: Genetec Institute, Inc.

STREET: 87 CambridgePark Drive

CITY: Cambridge

STATE: MA

COUNTRY: U.S.A.

ZIP: 02140

COMPUTER READABLE FORM:

MEDIUM TYPE: Floppy disk

COMPUTER: IBM PC compatible

OPERATING SYSTEM: PC-DOS/MS-DOS

SOFTWARE: PatentIn Release #1.0, Version #1.30

CURRENT APPLICATION DATA: US/09/745,763

Filing DATE: 18-Jun-2000

CLASSIFICATION: <Unknown>

ATTORNEY/AGENT INFORMATION:

NAME: Sprunger, Suzanne A.

REGISTRATION NUMBER: 41,323

TELECOMMUNICATION INFORMATION:

TELEPHONE: (617) 498-8284

TELEFAX: (617) 876-5851



INFORMATION FOR SEQ ID NO: 34:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 359 amino acids  
TYPE: amino acid  
STRANDEDNESS: <unknown>  
TOPOLOGY: linear  
MOLECULE TYPE: protein  
SEQUENCE DESCRIPTION: SEQ ID NO: 34:  
US-09-745-763-34

Query Match 100.0%; Score 1910; DB 9; Length 359;  
Best Local Similarity 100.0%; Pred. No. 1.2e-132;  
Matches 359; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 MAEPFLSEYQHQPOTSNCTGAAGAAVOBELNPERPPGAEEYVPEEDSRWQSAFAPQLGGRPG 60  
DB 1 MAEPFLSEYQHQPOTSNCTGAAGAAVOBELNPERPPGAEEYVPEEDSRWQSAFAPQLGGRPG 60  
QY 61 PEEGSLSEQPPPLQTOACPESSCLREGEKQNGDDSSAGDPPPAVEVPTPEAEILAQ 120  
DB 61 PEEGSLSEQPPPLQTOACPESSCLREGEKQNGDDSSAGDPPPAVEVPTPEAEILAQ 120  
QY 121 PCHDSEASKLGAAPAGGEEMGQQOQLGKKHRRRPSKKKHMKPYKLTWEKKKFE 180  
DB 121 PCHDSEASKLGAAPAGGEEMGQQOQLGKKHRRRPSKKKHMKPYKLTWEKKKFE 180  
QY 181 KOSLRASRIAEKFAKQVAPYNTTQFLMDHDQDEEPLKTGLYSKRAAKSDPISDD 240  
DB 181 KOSLRASRIAEKFAKQVAPYNTTQFLMDHDQDEEPLKTGLYSKRAAKSDPISDD 240  
QY 241 FMEEGEEDGSDGMDGSEFLQRFSEYERHYTESLQNMKSQELIKEYLEKCLSR 300  
DB 241 FMEEGEEDGSDGMDGSEFLQRFSEYERHYTESLQNMKSQELIKEYLEKCLSR 300  
QY 301 MEDENNRRLRESKRLGGDDARVARELELDRLRAENQLLTENELHQOERAPLSKFGD 359  
DB 301 MEDENNRRLRESKRLGGDDARVARELELDRLRAENQLLTENELHQOERAPLSKFGD 359

## RESULT 2

US-09-972-758-2  
Sequence 2, Application US/09972758  
Patent No. US20020160497A1  
GENERAL INFORMATION:  
APPLICANT: Case Western Reserve University  
APPLICANT: Montano, Monica  
APPLICANT: Wiltman, Bryan  
TITLE OF INVENTION: Suppressors of Human Breast Cancer Cell Growth  
FILE REFERENCE: 27708/04004  
CURRENT APPLICATION NUMBER: US/09/972, 758  
PRIOR FILING DATE: 2001-10-05  
PRIOR APPLICATION NUMBER: US 60/238,187  
PRIOR FILING DATE: 2000-10-05  
NUMBER OF SEQ ID NOS: 2  
SOFTWARE: PatentIn version 3.0  
SEQ ID NO 2  
LENGTH: 359  
TYPE: PRT  
ORGANISM: Homo sapiens  
US-09-972-758-2

Query Match 100.0%; Score 1910; DB 10; Length 359;  
Best Local Similarity 100.0%; Pred. No. 1.2e-132;  
Matches 359; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 MAEPFLSEYQHQPOTSNCTGAAGAAVOBELNPERPPGAEEYVPEEDSRWQSAFAPQLGGRPG 60  
DB 1 MAEPFLSEYQHQPOTSNCTGAAGAAVOBELNPERPPGAEEYVPEEDSRWQSAFAPQLGGRPG 60  
QY 61 PEEGSLSEQPPPLQTOACPESSCLREGEKQNGDDSSAGDPPPAVEVPTPEAEILAQ 120  
DB 61 PEEGSLSEQPPPLQTOACPESSCLREGEKQNGDDSSAGDPPPAVEVPTPEAEILAQ 120

QY 121 PCHDSEASKLGAAPAGGEEMGQQOQLGKKHRRRPSKKKHMKPYKLTWEKKKFE 180  
DB 121 PCHDSEASKLGAAPAGGEEMGQQOQLGKKHRRRPSKKKHMKPYKLTWEKKKFE 180  
QY 181 KOSLRASRIAEKFAKQVAPYNTTQFLMDHDQDEEPLKTGLYSKRAAKSDPISDD 240  
DB 181 KOSLRASRIAEKFAKQVAPYNTTQFLMDHDQDEEPLKTGLYSKRAAKSDPISDD 240  
QY 241 FMEEGEEDGSDGMDGSEFLQRFSEYERHYTESLQNMKSQELIKEYLEKCLSR 300  
DB 241 FMEEGEEDGSDGMDGSEFLQRFSEYERHYTESLQNMKSQELIKEYLEKCLSR 300  
QY 301 MEDENNRRLRESKRLGGDDARVARELELDRLRAENQLLTENELHQOERAPLSKFGD 359  
DB 301 MEDENNRRLRESKRLGGDDARVARELELDRLRAENQLLTENELHQOERAPLSKFGD 359

## RESULT 3

US-10-205-823-172  
Sequence 172, Application US/10205823  
Publication No. US20030108963A1  
GENERAL INFORMATION:  
APPLICANT: Schlegel, Robert  
APPLICANT: Monahan, John E.  
APPLICANT: Endege, Wilson O.  
APPLICANT: Ganavara, Manjula  
APPLICANT: Gorbacheva, Bella  
APPLICANT: Hoersch, Sebastian  
APPLICANT: Kamatkar, Shubhangi  
APPLICANT: Monsey, Angela M.  
APPLICANT: Glatc, Karen  
APPLICANT: Zhao, Xumel  
APPLICANT: Anderson, Dustin  
TITLE OF INVENTION: NOVEL GENES, COMPOSITIONS, KITS, AND  
METHODS FOR IDENTIFICATION, ASSESSMENT, PREVENTION, AND  
THERAPY OF PROSTATE CANCER  
FILE REFERENCE: MRI-044  
CURRENT APPLICATION NUMBER: US/10/205,823  
CURRENT FILING DATE: 2002-07-25  
PRIOR APPLICATION NUMBER: 60/307,982  
PRIOR FILING DATE: 2001-07-25  
PRIOR APPLICATION NUMBER: 60/314,356  
PRIOR FILING DATE: 2001-08-22  
PRIOR APPLICATION NUMBER: 60/325,020  
PRIOR FILING DATE: 2001-09-25  
PRIOR APPLICATION NUMBER: 60/341,746  
PRIOR FILING DATE: 2001-12-12  
PRIOR APPLICATION NUMBER: 60/362,158  
PRIOR FILING DATE: 2002-03-05  
NUMBER OF SEQ ID NOS: 455  
SOFTWARE: FastSeq for Windows Version 4.0  
SEQ ID NO 172  
LENGTH: 359  
TYPE: PRT  
ORGANISM: Homo sapiens  
US-10-205-823-172

Query Match 100.0%; Score 1910; DB 15; Length 359;  
Best Local Similarity 100.0%; Pred. No. 1.2e-132;  
Matches 359; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 MAEPFLSEYQHQPOTSNCTGAAGAAVOBELNPERPPGAEEYVPEEDSRWQSAFAPQLGGRPG 60  
DB 1 MAEPFLSEYQHQPOTSNCTGAAGAAVOBELNPERPPGAEEYVPEEDSRWQSAFAPQLGGRPG 60  
QY 61 PEEGSLSEQPPPLQTOACPESSCLREGEKQNGDDSSAGDPPPAVEVPTPEAEILAQ 120  
DB 61 PEEGSLSEQPPPLQTOACPESSCLREGEKQNGDDSSAGDPPPAVEVPTPEAEILAQ 120  
QY 121 PCHDSEASKLGAAPAGGEEMGQQOQLGKKHRRRPSKKKHMKPYKLTWEKKKFE 180  
DB 121 PCHDSEASKLGAAPAGGEEMGQQOQLGKKHRRRPSKKKHMKPYKLTWEKKKFE 180



QY 161 KOSLRASRI RAEMFAKGPVAPYNTTQFLMDHDOEBPDLKTGLYSKRAAASDDTSD 240  
 DB 161 KOSLRASRI RAEMFAKGPVAPYNTTQFLMDHDOEBPDLKTGLYSKRAAASDDTSD 240  
 QY 241 FMEEGEDGSDGMDGDSSEFLORDFSTYRHTYESIQNSKQELIKEYELEKCLSR 300  
 DB 241 FMEEGEDGSDGMDGDSSEFLORDFSTYRHTYESIQNSKQELIKEYELEKCLSR 300  
 QY 301 MEDENNRRLLESKRIGGDDARVRELELEDRLAENLQULTTENELHROOERAPLSKFGD 359  
 DB 301 MEDENNRRLLESKRIGGDDARVRELELEDRLAENLQULTTENELHROOERAPLSKFGD 359

## RESULT 4

US-10-094-749-2862  
 ; Sequence 2862, Application US/10094749  
 ; Publication No. US20030219741A1  
 ; GENERAL INFORMATION:  
 ; APPLICANT: ISOGAI, TAKAO  
 ; APPLICANT: SUGIYAMA, TOMOYASU  
 ; APPLICANT: OTSUKI, TETSUJI  
 ; APPLICANT: MAKAMATSU, AI  
 ; APPLICANT: SATO, HIROYUKI  
 ; APPLICANT: ISHII, SHIZUKO  
 ; APPLICANT: YAMAMOTO, JUN-ICHI  
 ; APPLICANT: ISONO, YUUKO  
 ; APPLICANT: HIO, YURI  
 ; APPLICANT: OTSUKA, KAORU  
 ; APPLICANT: NAGAI, KEIICHI  
 ; APPLICANT: IRIE, RYOTARO  
 ; APPLICANT: TAMECHIKA, ICHIRO  
 ; APPLICANT: SEKI, NAOHICO  
 ; APPLICANT: YOSHIKAWA, TSUTOMU  
 ; APPLICANT: OTSUKA, MOTOKYUKI  
 ; APPLICANT: NAGAHARI, KENJI  
 ; APPLICANT: MASUHO, YASUHIKO  
 ; TITLE OF INVENTION: NOVEL FULL-LENGTH CDNA  
 ; FILE REFERENCE: 084335/0160  
 ; CURRENT APPLICATION NUMBER: US/10/094,749  
 ; CURRENT FILING DATE: 2002-03-12  
 ; PRIOR APPLICATION NUMBER: 60/350,435  
 ; PRIOR FILING DATE: 2002-01-24  
 ; PRIOR APPLICATION NUMBER: JP 2001-328381  
 ; PRIOR FILING DATE: 2001-09-14  
 ; NUMBER OF SEQ ID NOS: 3381  
 ; SOFTWARE: Patentin Ver. 2.1  
 ; SEQ ID NO 2862  
 ; LENGTH: 286  
 ; TYPE: PRT  
 ; ORGANISM: Homo sapiens  
 ; US-10-094-749-2862

Query Match 28.6%; Score 545.5; DB 12; Length 286;  
 Best Local Similarity 43.8%; Pred. No. 2.3e-32;  
 Matches 128; Conservative 36; Mismatches 93; Indels 35; Gaps 6;  
 QY 73 PIOTQACPESSCLREBEKQNGD-----SAGDPPRAVEPTPEAEILLQPC-- 122  
 DB 5 PNOTACNASSPVALEEAKTSGAPSPQTPPERHDGSGSLPTPRMESHEDDLGAVGG 64  
 QY 123 --HDEASLGLGAPAAAGEEEMGQOQROLGKKHRRRPSKKKRWKPYULTWEEKKFPDE 180  
 DB 65 LGMNRSRPTQSPGGSAE-----AVLARKKHRRRPSKKKHWRYLELSNAEKGQRB 118  
 QY 181 KOSLRASRI RAEMFAKGPVAPYNTTQFLMDHDOEB--DKTGLYSKRAAASDDTSD 238  
 DB 119 ROSORASRYREEMFAKGPVAPYNTTQFLMDHDOEBPDLKTGLYSKRAAASDDTSD 173  
 QY 239 DPFMEEGEDGSDGMDGDSSEFLORDFSTYRHTYESIQNSKQELIKEYELEKCL 298  
 DB 174 -----EAGDSODGRGAHGFORKDSFETTERHTESIQNSKQELIKEYELEKCL 224  
 QY 299 SMWEDENNRRLLESKRIGGDDAR-VRELELEDRLAENLQULTTENELHROO 349

DB 225 SQAEETRRLOOLQACTGQOSCRQVEELAAEVOURLTENORLROSNQMNRE 276

## RESULT 5

US-09-976-740-44  
 ; Sequence 44, Application US/09976740  
 ; Publication No. US20020194633A1  
 ; GENERAL INFORMATION:  
 ; APPLICANT: Lees, Ann M.  
 ; APPLICANT: Lees, Robert S.  
 ; APPLICANT: Law, Simon W.  
 ; APPLICANT: Arjona, Anibal A.  
 ; TITLE OF INVENTION: NOVEL LOW DENSITY LIPOPROTEIN BINDING  
 ; TITLE OF INVENTION: PROTEINS AND THEIR USE IN DIAGNOSING AND TREATING  
 ; FILE REFERENCE: 10797-004001  
 ; CURRENT APPLICATION NUMBER: US/09/976,740  
 ; CURRENT FILING DATE: 2001-10-12  
 ; PRIOR APPLICATION NUMBER: 09/616,289  
 ; PRIOR FILING DATE: 2000-07-14  
 ; PRIOR APPLICATION NUMBER: US 08/979,608  
 ; PRIOR FILING DATE: 1997-11-26  
 ; PRIOR APPLICATION NUMBER: US 60/031,930  
 ; PRIOR FILING DATE: 1996-11-27  
 ; PRIOR APPLICATION NUMBER: US 60/048,547  
 ; PRIOR FILING DATE: 1997-06-03  
 ; NUMBER OF SEQ ID NOS: 53  
 ; SOFTWARE: FastSeq for Windows Version 4.0  
 ; SEQ ID NO 44  
 ; LENGTH: 546  
 ; TYPE: PRT  
 ; ORGANISM: Homo sapiens  
 ; US-09-976-740-44

Query Match 7.0%; Score 134.5; DB 10; Length 546;  
 Best Local Similarity 21.9%; Pred. No. 0.087;  
 Matches 86; Conservative 51; Mismatches 134; Indels 121; Gaps 14;

QY 20 GAAVQELINPERPGAEERVPEDSRWQSRFPOLGGRPGEGSGSLESQPPLOTQAC 79  
 DB 9 GAA--KQSNPSSPGQEPAGEGAEQERPPQAAAPV-----EAGPGSSQAP--RK 54  
 QY 80 PESSCLREGEKQNGD-----DSAGDPPRAVEPTPEAEILLQAP 121  
 DB 55 PEGAQRTAQSALRDVBEBSRLQEDLSTYCVNMGCPBEDAQGEF-----AEP 107  
 QY 122 CHDSBASKL-----GAPAAAGEEEMG-----QQQROLGKKHRRRPSKKKRW 163  
 DB 108 -EDAKSRITYAARNGEPPTPVNNGEKEPSKGDPTTEIRQSDVGDHRRRPSKKKRW 166  
 QY 164 W-----KPYULTWEEKKKPFPEKOSLASRIRAMFAKGPVAPYNTTQ 207  
 DB 167 GLGKEITLLMOTLNTLSTPEEBLALCKKYALBLEHNRNSQOMTLQKQ-----SQ 219  
 QY 208 FLMDHDOEBPDLKTGLYSKRAAASDDTSDPFMEEGEDGSDGMDGDSSEFLORDF 267  
 DB 220 LV-----QEKDHLRGHRSKAVLARSK-----LESCLREL 248  
 QY 268 SETYRHTYESIQNSKQELIKEYELEKCLSRMEDENNRRLLESKRIGGDDARVRELEL 327  
 DB 249 QR-----HNRSLKEBGVQRRAREEERKEVTSHPQVTLNDIQLOMEQHNHRSKRLROEWM 303  
 QY 328 ELDRLAENLQULTTENELHROOERAPLSKFGD 359  
 DB 304 EL-----AERLKLLEQYELREHIDKVFHKMD 331

## RESULT 6

US-10-616-187-44  
 ; Sequence 44, Application US/10616187  
 ; Publication No. US20040013668A1  
 ; GENERAL INFORMATION:

```

; APPLICANT: Lees, Ann M.
; APPLICANT: Lees, Robert S.
; APPLICANT: Law, Simon W.
; APPLICANT: Arjona, Anibal A.
; TITLE OF INVENTION: NOVEL LOW DENSITY LIPOPROTEIN BINDING
; TITLE OF INVENTION: PROTEINS AND THEIR USE IN DIAGNOSING AND TREATING
; FILE REFERENCE: 10797-004001
; CURRENT APPLICATION NUMBER: US/10/616,187
; CURRENT FILING DATE: 2003-07-09
; PRIOR APPLICATION NUMBER: US/09/616,289
; PRIOR FILING DATE: 2000-07-14
; PRIOR APPLICATION NUMBER: US 09/517,849
; PRIOR FILING DATE: 2000-03-02
; PRIOR APPLICATION NUMBER: US 08/979,608
; PRIOR FILING DATE: 1997-11-26
; PRIOR APPLICATION NUMBER: US 60/031,930
; PRIOR FILING DATE: 1996-11-27
; PRIOR APPLICATION NUMBER: US 60/048,547
; PRIOR FILING DATE: 1997-06-03
; NUMBER OF SEQ ID NOS: 53
; SOFTWARE: FaastSeq for Windows Version 4.0
; SEQ ID NO 44
; LENGTH: 546
; TYPE: PRT
; ORGANISM: Homo sapiens
US-10-616-187-44

```

Query Match 7.0%; Score 134.5; DB 12; Length 546;

Best Local Similarity 21.9%; Pred. No. 0.087; Matches 86; Conservative 51; Mismatches 134; Indels 121; Gaps 14;

```

QY 20 GAAVQELINPERPPAEERVPEDSRWQRAFPOLGGRPGEGSGLESQPPPLQTOAC 79
   |||::|||::|||::|||::|||::|||::|||::|||::|||::|||::|||::|||
DB 9 GAA---KQSNPKSPQGPAGPEGAGQERPSQAAPAV-----EABPGSSQAP-----RK 54
QY 80 PSSSCIREGKQND-----SSAGGDPFPAPAEVETPTEAELLAP 121
   |||::|||::|||::|||::|||::|||::|||::|||::|||::|||::|||::|||
DB 55 PEGAQARTQSGALRDVSEELSRQLEDILSTYCVNNQGPBGDGAQGP-----ABP 107
QY 122 CHDSKSKL-----GAPAGGEEWNG-----QQOROLGKKHRRRPSKKKH 163
   |||::|||::|||::|||::|||::|||::|||::|||::|||::|||::|||::|||
DB 108 -EDAERSRTYVANGPEPTPVNNGEKESKDPNTEIRQSDVEVDGRHRRRQEKKKAK 166
QY 164 W-----KPYKLTWEKKKFKDEKQSLRASRIAEFAKQVPAPYNTTQ 207
   |||::|||::|||::|||::|||::|||::|||::|||::|||::|||::|||::|||
DB 167 GKGKETLLMQTLNTLSTPEEKLAALCKKYAELLEHRNSQKMKLQKKQ-----SQ 219
QY 208 FLMDHDQEPPLKTGLYKRAAAKSDDTSDDDFMEEGGEGGSDMGDSSEFLQDPF 267
   |||::|||::|||::|||::|||::|||::|||::|||::|||::|||::|||::|||
DB 220 LV-----QEKDLRGESHKAVLARSK-----LESICREL 248
QY 268 SETERYHTESLQNNKQKELIKYLELEKCLSRMEDENNRLRLESKRLGGDDARVLEL 327
   |||::|||::|||::|||::|||::|||::|||::|||::|||::|||::|||::|||
DB 249 QR-----HNRSLKEGVQARAREEERKEVTSHPQVTLNDIQLOMEQHNRNSKLRQENM 303
QY 328 ELDRLEAENLQLTENELHROQERAPLSKFGD 359
   |||::|||::|||::|||::|||::|||::|||::|||::|||::|||::|||::|||
DB 304 EL-----AERLKKLIEQYELREEHIDKVFQHKD 331

```

```

RESULT 7
US-10-023-529-44
; Sequence 44, Application US/10023529
; Publication No. US20020129386A1
; GENERAL INFORMATION:
; APPLICANT: Lees, Ann M.
; APPLICANT: Lees, Robert S.
; APPLICANT: Law, Simon W.
; APPLICANT: Arjona, Anibal A.
; TITLE OF INVENTION: NOVEL LOW DENSITY LIPOPROTEIN BINDING
; TITLE OF INVENTION: PROTEINS AND THEIR USE IN DIAGNOSING AND TREATING
; TITLE OF INVENTION: ATHEROSCLEROSIS

```

```

; FILE REFERENCE: 10797-004001
; CURRENT APPLICATION NUMBER: US/10/023,529
; CURRENT FILING DATE: 2001-12-17
; PRIOR APPLICATION NUMBER: 09/616,289
; PRIOR FILING DATE: 2000-07-14
; PRIOR APPLICATION NUMBER: US 09/517,849
; PRIOR FILING DATE: 2000-03-02
; PRIOR APPLICATION NUMBER: US 08/979,608
; PRIOR FILING DATE: 1997-11-26
; PRIOR APPLICATION NUMBER: US 60/031,930
; PRIOR FILING DATE: 1996-11-27
; PRIOR APPLICATION NUMBER: US 60/048,547
; PRIOR FILING DATE: 1997-06-03
; NUMBER OF SEQ ID NOS: 53
; SOFTWARE: FaastSeq for Windows Version 4.0
; SEQ ID NO 44
; LENGTH: 546
; TYPE: PRT
; ORGANISM: Homo sapiens
US-10-023-529-44

```

Query Match 7.0%; Score 134.5; DB 14; Length 546;

Best Local Similarity 21.9%; Pred. No. 0.087; Matches 86; Conservative 51; Mismatches 134; Indels 121; Gaps 14;

```

QY 20 GAAVQELINPERPPAEERVPEDSRWQRAFPOLGGRPGEGSGLESQPPPLQTOAC 79
   |||::|||::|||::|||::|||::|||::|||::|||::|||::|||::|||::|||
DB 9 GAA---KQSNPKSPQGPAGPEGAGQERPSQAAPAV-----EABPGSSQAP-----RK 54
QY 80 PSSSCIREGKQND-----SSAGGDPFPAPAEVETPTEAELLAP 121
   |||::|||::|||::|||::|||::|||::|||::|||::|||::|||::|||::|||
DB 55 PEGAQARTQSGALRDVSEELSRQLEDILSTYCVNNQGPBGDGAQGP-----ABP 107
QY 122 CHDSKSKL-----GAPAGGEEWNG-----QQOROLGKKHRRRPSKKKH 163
   |||::|||::|||::|||::|||::|||::|||::|||::|||::|||::|||::|||
DB 108 -EDAERSRTYVANGPEPTPVNNGEKESKDPNTEIRQSDVEVDGRHRRRQEKKKAK 166
QY 164 W-----KPYKLTWEKKKFKDEKQSLRASRIAEFAKQVPAPYNTTQ 207
   |||::|||::|||::|||::|||::|||::|||::|||::|||::|||::|||::|||
DB 167 GKGKETLLMQTLNTLSTPEEKLAALCKKYAELLEHRNSQKMKLQKKQ-----SQ 219
QY 208 FLMDHDQEPPLKTGLYKRAAAKSDDTSDDDFMEEGGEGGSDMGDSSEFLQDPF 267
   |||::|||::|||::|||::|||::|||::|||::|||::|||::|||::|||::|||
DB 220 LV-----QEKDLRGESHKAVLARSK-----LESICREL 248
QY 268 SETERYHTESLQNNKQKELIKYLELEKCLSRMEDENNRLRLESKRLGGDDARVLEL 327
   |||::|||::|||::|||::|||::|||::|||::|||::|||::|||::|||::|||
DB 249 QR-----HNRSLKEGVQARAREEERKEVTSHPQVTLNDIQLOMEQHNRNSKLRQENM 303
QY 328 ELDRLEAENLQLTENELHROQERAPLSKFGD 359
   |||::|||::|||::|||::|||::|||::|||::|||::|||::|||::|||::|||
DB 304 EL-----AERLKKLIEQYELREEHIDKVFQHKD 331

```

```

RESULT 8
US-10-023-523-44
; Sequence 44, Application US/10023523
; Publication No. US20020152485A1
; GENERAL INFORMATION:
; APPLICANT: Lees, Ann M.
; APPLICANT: Lees, Robert S.
; APPLICANT: Law, Simon W.
; APPLICANT: Arjona, Anibal A.
; TITLE OF INVENTION: NOVEL LOW DENSITY LIPOPROTEIN BINDING
; TITLE OF INVENTION: PROTEINS AND THEIR USE IN DIAGNOSING AND TREATING
; FILE REFERENCE: 10797-004001
; CURRENT APPLICATION NUMBER: US/10/023,523
; CURRENT FILING DATE: 2001-12-17
; PRIOR APPLICATION NUMBER: US/09/616,289
; PRIOR FILING DATE: 2000-07-14
; PRIOR APPLICATION NUMBER: US 09/517,849
; PRIOR FILING DATE: 2000-03-02

```

```

; PRIOR APPLICATION NUMBER: US 08/979,608
; PRIOR FILING DATE: 1997-11-26
; PRIOR APPLICATION NUMBER: US 60/031,930
; PRIOR FILING DATE: 1996-11-27
; PRIOR APPLICATION NUMBER: US 60/048,547
; PRIOR FILING DATE: 1997-06-03
; NUMBER OF SEQ ID NOS: 53
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 44
; LENGTH: 546
; TYPE: PR1
; ORGANISM: Homo sapiens
; US-10-023-523-44

```

```

Query Match      7.0%; Score 134.5; DB 14; Length 546;
Best Local Similarity 21.9%; Pred. No. 0.087;
Matches 86; Conservative 51; Mismatches 134; Indels 121; Gaps 14;

```

```

QY 20 GAAVAVELNPERPPGAERVPEDSRWQSRAPFOLGRRPGEGSLGQPPPLQTOAC 79
DB 9 GAAVAVELNPERPPGAERVPEDSRWQSRAPFOLGRRPGEGSLGQPPPLQTOAC 79
QY 80 PESSCLREKGGQND-----DSAGDPFPPPAVEPTPAELLAP 121
DB 55 PEGAQARTAQSGALRDVSEELSRQLEDILSTYCVNNQSGPGEDEAGQGP-----ABP 107
QY 122 CHDSEASKL-----GAPAGGEEWNG-----QQOROLGKKRRRPPSKKRRH 163
DB 108 -EDAEKSRITYVANRGEPEPPVNVNGEKEPKGDPNTEIRIQSDPEVGRDRRRPOEKKAK 166
QY 164 W-----KPYKLTWEKKKKFDEKOSLRASRIRAEPAKQOPVAPYNTTQ 207
DB 167 GLGKEITLMQTLNTLSTPEEKLAALCKYAEELLEHRNQQKMKLLQKKQ-----SQ 219
QY 208 FLMDHDOEPDLKTYGYSRAAKSDDTSDDDFMEEGSGEDGSGDGGSEPLQDRF 267
DB 220 LV-----QEKDHLRGHSHKAVLARSK-----LESICREL 248
QY 268 SEFYERYHTSLQMSKQELIKYELKELKSLMEDENNRLRLSKRLGGDDARVLEL 327
DB 249 QR-----HNRSLKEGVORAREEKEKKEVTSHPQVTLNDIQLOMEQHNRNKKLQENM 303
QY 328 ELDRLEAENLQLTENELHROERAPLSKFGD 359
DB 304 EL-----AERLKLILEYELAEHNDKVPFKKD 331

```

```

RESULT 9
US-09-216-393-81
; Sequence 81, Application US/09216393
; Patent No. US20010014447A1
; GENERAL INFORMATION:
; APPLICANT: Milhausen, Michael James
; TITLE OF INVENTION: TOXOPLASMA GONDII PROTEINS, NUCLEIC ACID MOLECULES, AND
; FILE REFERENCE: TX-1-C2
; CURRENT APPLICATION NUMBER: US/09/216,393
; CURRENT FILING DATE: 1998-12-18
; EARLIER APPLICATION NUMBER: 08/994,825
; EARLIER FILING DATE: 1997-12-19
; NUMBER OF SEQ ID NOS: 364
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 81
; LENGTH: 611
; TYPE: PR1
; ORGANISM: Toxoplasma gondii
; US-09-216-393-81

```

```

Query Match      6.8%; Score 130; DB 9; Length 611;
Best Local Similarity 22.2%; Pred. No. 0.21;
Matches 82; Conservative 58; Mismatches 155; Indels 74; Gaps 15;
QY 13 POTSNGTGAAYVEELN-PPPPGAERVP-----EEDSRWQSRAPF-----QGGRRGPE 62

```

```

DB 189 PFFSKSDVCCSPQARLSLPEOSLGSSPSPISVTNDVVALFDSASAPLHAGELSSLPQAV 248
QY 63 GGGSLGSP-----PLQTOACPESSCLREGEKGQNGDDSSAGDPPPAVEPTPAELL 118
DB 249 SASERLLTAPAITGSSASACLSVSC-----GPEMSPTADTTR- 287
QY 119 AOPCHDSEASKLGAAPAAGE-----EEMGOOQOLGKKRRRPPSKKRRMKPYKLTWE 174
DB 288 -----HDAEERERRRRAEEERERERERERERERERERERERERERERERERERER 339
QY 175 K--KKFDEKQSLRA-----SRIAEWFAKQOPVAPYNTTQFLMDHDOEPDLKTYGK 227
DB 340 KAROEEDERERRRRAVEEERAROREEERERRRRAVEEERARQ-----REEEERERRRVE 395
QY 228 RAAKSDDTSDDDFMEE-----GGEEDGSGDGMGDSGLQDSFLODFSEITYE--RYHTESLQ 281
DB 396 AROREEERERRRRAVEEERAROREEERERRRRAVEEERARQ-----REEEERERRRVE 455
QY 282 MSKQELIKEYLELEKLSRMEDENNRLRLSKRLGGDDARVLELDRLEAENLQLLT 341
DB 456 REEEERERRRRAVEEERAROREEERERRRRAVEEERARQ-----ERREQEEERERRRVE 505
QY 342 ENELHROQE 350
DB 506 EKEREROQE 514

```

```

RESULT 10
US-10-321-856-81
; Sequence 81, Application US/10321856
; Publication No. US20030194393A1
; GENERAL INFORMATION:
; APPLICANT: Milhausen, Michael James
; TITLE OF INVENTION: TOXOPLASMA GONDII PROTEINS, NUCLEIC ACID MOLECULES, AND USES THERE
; FILE REFERENCE: TX-1-C2-1
; CURRENT APPLICATION NUMBER: US/10/321,856
; CURRENT FILING DATE: 2002-12-17
; PRIOR APPLICATION NUMBER: 09/216,393
; PRIOR FILING DATE: 1998-12-18
; PRIOR APPLICATION NUMBER: 08/994,825
; PRIOR FILING DATE: 1997-12-19
; NUMBER OF SEQ ID NOS: 366
; SOFTWARE: PatentIn version 3.1
; SEQ ID NO 81
; LENGTH: 611
; TYPE: PR1
; ORGANISM: Toxoplasma gondii
; US-10-321-856-81

```

```

Query Match      6.8%; Score 130; DB 12; Length 611;
Best Local Similarity 22.2%; Pred. No. 0.21;
Matches 82; Conservative 58; Mismatches 155; Indels 74; Gaps 15;
QY 13 POTSNGTGAAYVEELN-PPPPGAERVP-----EEDSRWQSRAPF-----QGGRRGPE 62
DB 189 PFFSKSDVCCSPQARLSLPEOSLGSSPSPISVTNDVVALFDSASAPLHAGELSSLPQAV 248
QY 63 GGGSLGSP-----PLQTOACPESSCLREGEKGQNGDDSSAGDPPPAVEPTPAELL 118
DB 249 SASERLLTAPAITGSSASACLSVSC-----GPEMSPTADTTR- 287
QY 119 AOPCHDSEASKLGAAPAAGE-----EEMGOOQOLGKKRRRPPSKKRRMKPYKLTWE 174
DB 288 -----HDAEERERRRRAEEERERERERERERERERERERERERERERERERERER 339
QY 175 K--KKFDEKQSLRA-----SRIAEWFAKQOPVAPYNTTQFLMDHDOEPDLKTYGK 227
DB 340 KAROEEDERERRRRAVEEERAROREEERERRRRAVEEERARQ-----REEEERERRRVE 395
QY 228 RAAKSDDTSDDDFMEE-----GGEEDGSGDGMGDSGLQDSFLODFSEITYE--RYHTESLQ 281
DB 396 AROREEERERRRRAVEEERAROREEERERRRRAVEEERARQ-----ERREQEEERERRRVE 455

```

QY 282 MSKQELIKYLELEKLSRMEDENNRRLLESKRIGDADARVLELELDRLAENIQLLT 341  
 Db 456 REEERERRRRVEEKEERERQEEERERRRRVEEKE-----EREROEEERERRRVEE----- 505  
 QY 342 ENELHROOE 350  
 Db 506 EKEREROE 514

RESULT 11  
 US-10-108-260A-4161  
 ; Sequence 4161, Application US/10108260A  
 ; Publication No. US20040005560A1  
 ; GENERAL INFORMATION:  
 ; APPLICANT: HELIX RESEARCH INSTITUTE  
 ; TITLE OF INVENTION: No. US20040005560A1el full length cdna  
 ; FILE REFERENCE: HI-A0106  
 ; CURRENT APPLICATION NUMBER: US/10/108,260A  
 ; NUMBER OF SEQ ID NOS: 5458  
 ; SOFTWARE: PatentIn Ver. 2.1  
 ; SEQ ID NO 4161  
 ; LENGTH: 804  
 ; TYPE: PRT  
 ; ORGANISM: Homo sapiens  
 ; US-10-108-260A-4161

Query Match 6.7%; Score 128.5; DB 12; Length 804;  
 Best Local Similarity 21.3%; Pred. No. 0.38; Indels 165; Gaps 21;  
 Matches 96; Conservative 52; Mismatches 137;  
 QY 5 FLSEYHQPQTSNCTGAAGVQEBLNPERRPGAE--ERV--PER-----DSRWQ 48  
 Db 173 FSSIRRH--RKSVTGA-----EGSEPGAKGPEVRARPHVHSARQVCFEETFQ 222  
 QY 49 S-----RAFPOLGGRPGEGSGLESOPPLQTOAC-----PSSGC 84  
 Db 223 APRKENANPQ--DAPGPKVSPTEPPSPATEKMACKDPKPMKACASAHVQPKPAPASS 280  
 QY 85 LRGEKQNGDDSSAGGDFPPPAEV--EP-----TPBAELLAQPCDSE 126  
 Db 281 LEEPPHSETEKVVAGVNPNGVDPPLSLFGDVTSLKSPSLTGCGGIIABQMDSM 340  
 QY 127 ASKLGAPAGG-----EEWQOQROLGKKKRR 155  
 Db 341 TDSM--ASGGQANRDTGRSSCLVTVYGGGEMALPDDDDDEEBEEBEVELEBER-- 394  
 QY 156 RPSKKRKHMKPYKLTWEKKKDEKQSLASRIAPAMFAKQGVAPYNTQ-----FL 209  
 Db 395 -----EEVKEEBEDDLEYLWETAQMYPRPMNMLGYHPTTSPGHNGY 437  
 QY 210 MDDHDEEPDLKTG--LYSKRAAKSDTSDDDFMER--GSEEDGSGDGMGSGSEFLOR 265  
 Db 438 LDPVRSYPLGALPGBELTTPQSDQESAPNSDEGYDTPGFEEDSGEALGLVRDCLPR 497  
 QY 266 DF--SETVRYH--TESLQNM-----SKQELIKYLELEKLS-----EMEDE 304  
 Db 498 DVSYGDALVFYEPDSDLENSPPGDDCLVLDHGRSSMFPPLNFEPFLSSRRPGAMETE 557  
 QY 305 NNRL-----RLESKRIGDADARVRE 324  
 Db 558 EERLVTTIQQLYWELELRQLEAERARE 587

RESULT 12  
 US-09-824-574-7  
 ; Sequence 7, Application US/09824574  
 ; Publication No. US20030077800A1  
 ; GENERAL INFORMATION:  
 ; APPLICANT: Rouleau, Natalie  
 ; APPLICANT: Moilanen, Anu-Maarit  
 ; APPLICANT: Palvimio, Jorma J.

; APPLICANT: Time, Olli A,  
 ; TITLE OF INVENTION: Arip4 Gene and Protein  
 ; FILE REFERENCE: 2630-109  
 ; CURRENT APPLICATION NUMBER: US/09/824,574  
 ; CURRENT FILING DATE: 2001-04-03  
 ; NUMBER OF SEQ ID NOS: 7  
 ; SOFTWARE: PatentIn version 3.0  
 ; SEQ ID NO 7  
 ; LENGTH: 2476  
 ; TYPE: PRT  
 ; ORGANISM: Mus musculus  
 ; US-09-824-574-7

Query Match 6.7%; Score 128; DB 11; Length 2476;  
 Best Local Similarity 19.7%; Pred. No. 1.6;  
 Matches 78; Conservative 52; Mismatches 130; Indels 136; Gaps 13;  
 QY 43 EDSRWSSRAFPOLGGRPGEGE--GSLSSOPPLQTOACPFSSCRLREBEKQNGDSSA 99  
 Db 1142 EDKKQKKQRTSAKKKTGNTKEKRNSLRATPRKQVDITSSSDIGDDQNSAGESSD 1201  
 QY 100 GGDPPP-----PAVEP-----TPE-----AELIA 119  
 Db 1202 EOKIKVTENLVLPSTHFGCQSSGDEALSKVPATVDDDDDDNDPENRIAKKYLLEIRA 1261  
 QY 120 QPCHDSEASKLGAAPAGSEEWQOQ-----RQIGKKKRRRPSKK- 161  
 Db 1262 NLSDEDSGSDDEPDGCKKRIQGEESPADDELRRQLAVNQVNSSESDSSESKKP 1321  
 QY 162 --RHMKPYKLTWEKKKDEKQSLASRIAPAMFAKQGVAPYNTQGLMDHDEED 219  
 Db 1322 RYRHLRLRHKLTDSESGEEK--PTPKHKEKGR----- 1356  
 QY 220 LKTGLYSKRAAKSDTSDDDFMEEGEGEDGSDGSGSEFLQDFSETYERYHTSL 279  
 Db 1357 -----NRKYSSESEDTDFQESVSEVSE-----SDEQRPRTSAKKALEEN 1402  
 QY 280 QNMSKQELIKYLELEKLSR-----MEDENNRRLLESKRL 315  
 Db 1403 QRSYKQKKRRRIKVQEDSSSENKSHSEBKKEGDEDEDEDEDEDEND--DSKSP 1458  
 QY 316 GGDADARVLELELDRLAENIQLLTENLHQER 351  
 Db 1459 GKGRKKIRKI-LKDDLRTE-----TONALKEEER 1488

RESULT 13  
 US-10-161-051-193  
 ; Sequence 193, Application US/10161051  
 ; Publication No. US20030152945A1  
 ; GENERAL INFORMATION:  
 ; APPLICANT: Peter Deak  
 ; APPLICANT: David M Glover  
 ; APPLICANT: Carol Midgley  
 ; TITLE OF INVENTION: Cell cycle progression proteins  
 ; FILE REFERENCE: CCI-021CP  
 ; CURRENT APPLICATION NUMBER: US/10/161,051  
 ; CURRENT FILING DATE: 2002-05-30  
 ; PRIOR APPLICATION NUMBER: GB 0007268.6  
 ; NUMBER OF SEQ ID NOS: 194  
 ; SOFTWARE: PatentIn Ver. 2.1  
 ; SEQ ID NO 193  
 ; LENGTH: 722  
 ; TYPE: PRT  
 ; ORGANISM: Drosophila melanogaster  
 ; US-10-161-051-193

Query Match 6.6%; Score 126.5; DB 12; Length 722;  
 Best Local Similarity 22.5%; Pred. No. 0.47;  
 Matches 84; Conservative 47; Mismatches 149; Indels 93; Gaps 16;  
 QY 12 QPQTSNC--TGAAGVQEBLNPERRPGAEERVP--EEDSRWQSAFPOLGGRPGEGSGLE 68

```

Db      324 EPEBSJCLBEGDEEBEDADEDLDEDEEDPSEBEDKRRS-----GKSGGAGRGASAR 377
Qy      69 SQP--PILOT-----OACPESSCLREGE-----KQONGDSSAGGDFPPPAVEPTPEA 115
Db      378 NSGRPRRATAGKMAHYVDFSSSDSEQVAVPKRRRNDSSGSDYNSAN-----430
Qy      116 ELLAOPCHSEASKLGAIPAAGEEWGQOOLGKKKRRRPSKKRHWKPYKLTWEK 175
Db      431 -----SPDGGRGGAAGAAGKRVPSRGGRPARKSRRNSDSE-----EBE 473
Qy      176 KKFDEKOSLRASIRAEWFAKGPVAPYNTTQFLMDHDOEBDLTGLYSK-RAAA---231
Db      474 SEVSDADSDVPKRRKRGSGVGRGRPAAP-----ASAGRRGRGRGAASRK 516
Qy      232 KSDDTGDDDMERGESEEDSGSDMGCGDSEFLQDPEST-----YERYHTESTLQMSKQ 285
Db      517 RKDSDSEDEVSDEEBEDVD--PASDQSEVCKNLISSTWCFTKTPITQBERPKSKK 575
Qy      286 ELIKEYLELEKCLSRMEDENNRLLESKRLGDDARVRELELDRLRAENLQLTENEL 345
Db      576 PITPA-----KNSXANNKSKPAGKADSRKSKKSSSE---EDDVDDKDE- 618
Qy      346 HQQERAPLSKFG 358
Db      619 --SDEDEPLTKG 629

```

```

RESULT 14
US-10-259-194A-320
; Sequence 320, Application US/10259194A
; Publication No. US20040010815A1
; GENERAL INFORMATION:
; APPLICANT: Lange, Markus B.
; APPLICANT: Ghassemlan, Majid
; APPLICANT: Briggs, Steven P.
; APPLICANT: Cooper, Bret
; APPLICANT: Glazebrook, Jane
; APPLICANT: Goff, Stephen A.
; APPLICANT: Katagiri, Rumiya
; APPLICANT: Kreps, Joel
; APPLICANT: Moughamer, Todd
; APPLICANT: Provart, Nicholas
; APPLICANT: Ricker, Darrell
; APPLICANT: Zhu, Tong
; TITLE OF INVENTION: IDENTIFICATION AND CHARACTERIZATION OF PLANT GENES
; FILE REFERENCE: 70029-NP
; CURRENT APPLICATION NUMBER: US/10/259,194A
; PRIOR FILING DATE: 2003-01-07
; PRIOR APPLICATION NUMBER: US 60/325,277
; PRIOR FILING DATE: 2001-09-26
; PRIOR APPLICATION NUMBER: US 60/370,743
; PRIOR FILING DATE: 2002-04-04
; PRIOR APPLICATION NUMBER: US 60/370,620
; NUMBER OF SEQ ID NOS: 662
; SOFTWARE: Patentlist.pl version 3.0.4 (C) 2001 Syngenta
; SEQ ID NO 320
; LENGTH: 323
; TYPE: PRT
; ORGANISM: Oryza sativa
US-10-259-194A-320

```

```

Query Match      6.5%; Score 125; DB 12; Length 323;
Best Local Similarity 21.6%; Pred. No. 0.23;
Matches 66; Conservative 60; Mismatches 130; Indels 50; Gaps 13;

```

```

Qy      64 EGSIESPPPLQACPESSCLREGEKQ-ONGDSSAG-----GDFPPPAVEPTPEAKEL 118
Db      7 QNSIKGTGDKLPVTRTSSGEEGSRAREAKRANSRRKIKDF-----SADLE 56
Qy      119 AOPCHSEASKLGAIPA--GSEEWGQOOLGKKKRRRPSKKRHWKPYKLTWEK 177

```

```

Db      57 LKKAHDLSPEKSPSQSHSKETRRKONQLESSESEDEDEGRITK-----TVDSPPD 109
Qy      178 FDEKOSLRASIRAEWFAKGPVAPYNTTQFLMDHDOEE--PDLTKGLYSKRAAKSD 235
Db      110 SQQKH-TPSR-----VGMHNS--YKDGNNSSDALKGLRDGMAKKYPKXID 155
Qy      236 TSDDD---FMEEGGEEDGSD-GMGDSEFLQDPESTERYHTESTLQMSKOELE 290
Db      156 DSESEDSPPRXDKRAHGNNDSSGSGESGKHSERKHKKSRKRRYDSSD 215
Qy      291 YLELE---KCLSRMEDENNRLLESKRLGDDARVRELELDRLR---AENLQLTENE 344
Db      216 ESNSESDDESRRRKEERLKEERRRRERHRKADRSKJLKRAETVDMAADLE 275
Qy      345 LHRQOE 350
Db      276 KDRSD 281

```

```

RESULT 15
US-10-374-780A-2854
; Sequence 2854, Application US/10374780A
; Publication No. US20040019927A1
; GENERAL INFORMATION:
; APPLICANT: Sherman, Bradley K.
; APPLICANT: Riechmann, Jose Luis
; APPLICANT: Jiang, Cai-Zhong
; APPLICANT: Heard, Jacqueline E.
; APPLICANT: Haake, Volker
; APPLICANT: Creelman, Robert A.
; APPLICANT: Ratcliffe, Oliver
; APPLICANT: Adam, Luc J.
; APPLICANT: Reuber, T. Lynne
; APPLICANT: Keddie, James
; APPLICANT: Brown, Pierre E.
; APPLICANT: Pilgrim, Marsha L.
; APPLICANT: Dubell III, Arnold T.
; APPLICANT: Pineda, Omaira
; APPLICANT: Yu, Guo-Liang
; TITLE OF INVENTION: POLYNUCLEOTIDES AND POLYPEPTIDES IN PLANTS
; FILE REFERENCE: MB1-0047 CIP
; CURRENT APPLICATION NUMBER: US/10/374,780A
; PRIOR FILING DATE: 2003-02-25
; PRIOR APPLICATION NUMBER: 09/837,944
; PRIOR FILING DATE: 2001-04-18
; PRIOR APPLICATION NUMBER: 60/310,847
; PRIOR FILING DATE: 2001-08-09
; PRIOR APPLICATION NUMBER: 09/934,455
; PRIOR FILING DATE: 2001-08-22
; PRIOR APPLICATION NUMBER: 60/336,049
; PRIOR FILING DATE: 2001-11-19
; PRIOR APPLICATION NUMBER: 60/338,692
; PRIOR FILING DATE: 2001-12-11
; PRIOR APPLICATION NUMBER: 10/171,468
; PRIOR FILING DATE: 2002-06-14
; PRIOR APPLICATION NUMBER: 10/225,066
; PRIOR FILING DATE: 2002-08-09
; PRIOR APPLICATION NUMBER: 10/225,067
; PRIOR FILING DATE: 2002-08-09
; PRIOR APPLICATION NUMBER: 10/225,068
; PRIOR FILING DATE: 2002-08-09
; NUMBER OF SEQ ID NOS: 2906
; SOFTWARE: PatentIn version 3.2
; SEQ ID NO 2854
; LENGTH: 383
; TYPE: PRT
; ORGANISM: Arabidopsis thaliana
; FEATURE:
; OTHER INFORMATION: G2373
US-10-374-780A-2854

```

```

Query Match      6.5%; Score 125; DB 12; Length 383;
Best Local Similarity 20.6%; Pred. No. 0.29;

```

